

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
24 July 2003 (24.07.2003)

PCT

(10) International Publication Number  
WO 03/060149 A2(51) International Patent Classification<sup>7</sup>: C12Q

US 60/396,407 (CIP)

(21) International Application Number: PCT/US03/00252

Filed on Not furnished

US 60/415,115 (CIP)

Filed on Not furnished

(22) International Filing Date: 6 January 2003 (06.01.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/345,222	4 January 2002 (04.01.2002)	US
60/348,693	14 January 2002 (14.01.2002)	US
60/349,182	16 January 2002 (16.01.2002)	US
60/349,733	17 January 2002 (17.01.2002)	US
60/350,263	18 January 2002 (18.01.2002)	US
60/351,977	24 January 2002 (24.01.2002)	US
60/383,758	28 May 2002 (28.05.2002)	US
60/385,969	5 June 2002 (05.06.2002)	US
60/387,834	11 June 2002 (11.06.2002)	US
60/396,407	17 July 2002 (17.07.2002)	US
60/415,115	30 September 2002 (30.09.2002)	US
10/336,603	3 January 2003 (03.01.2003)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	10/336,603 (CIP)
Filed on	Not furnished
US	60/345,222 (CIP)
Filed on	Not furnished
US	60/348,693 (CIP)
Filed on	Not furnished
US	60/349,182 (CIP)
Filed on	Not furnished
US	60/349,733 (CIP)
Filed on	Not furnished
US	60/350,263 (CIP)
Filed on	Not furnished
US	60/351,977 (CIP)
Filed on	Not furnished
US	60/383,758 (CIP)
Filed on	Not furnished
US	60/385,969 (CIP)
Filed on	Not furnished
US	60/387,834 (CIP)
Filed on	Not furnished

(71) Applicant (for all designated States except US): CURAGEN CORPORATION [US/US]; 555 Long Wharf Drive, 11th floor, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): GROSSE, William, M. [US/US]; 15 Rice Terrace Road, Apartment C, Branford, CT 06405 (US). ALSOBROOK, II, John, P. [US/US]; 60 Lake Drive, Madison, CT 06443 (US). ANDERSON, David, W. [US/US]; 85 Montoya Drive, Branford, CT 06405 (US). BURGESS, Catherine, E. [US/US]; 90 Carriage Hill Drive, Wethersfield, CT 06109 (US). EDINGER, Shlomit, R. [US/US]; 766 Edgewood Avenue, New Haven, CT 06515 (US). ELLERMAN, Karen [US/US]; 87 Montoya Drive, Branford, CT 06405 (US). FURTAK, Katarzyna [PL/US]; 31 Park Place, Ansonia, CT 06401 (US). GANGOLLI, Esha, A. [IN/US]; 31 Strawberry Hill Road, Madison, CT 06443 (US). GERLACH, Valerie, L. [US/US]; 18 Rock Pasture Road, Branford, CT 06405 (US). GILBERT, Jennifer, A. [US/US]; 343 Horsepond Road, Madison, CT 06443 (US). GUNTHER, Erik [US/US]; 34 Byan Road, Branford, CT 06405 (US). GORMAN, Linda [US/US]; 329 Monticello Drive, Branford, CT 06405 (US). GUO, Xiaojia [CN/US]; 713 Robert Frost Drive, Branford, CT 06405 (US). JI, Weizhen [CN/US]; 375 Old Rock Road, Branford, CT 06405 (US). LI, Li [CN/US]; 56 Jerimoth Drive, Branford, CT 06405 (US). MILLER, Charles, E. [US/US]; 98 Saddle Hill Drive, Guilford, CT 06437 (US). PADIGARU, Muralidhara [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). PATTURAJAN, Meera [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, CT 06405 (US). RASTELLI, Luca [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). MACDOUGALL, John, R. [CA/US]; 117 Russell Street, Hamden, CT 06517 (US). MISHRA, Vishnu, S. [IN/US]; 3832 SouthWest 93rd Terrace, Gainesville, FL 32608 (US). SMITHSON, Glenda [US/US]; 125 Michael Drive, Guilford, CT 06435 (US). SPYTEK, Kimberly, A. [US/US]; 28 Court Street, Number 1, New Haven, CT 06511 (US). STONE, David, J. [US/US]; 223 Whitehorn Drive, Guilford, CT

[Continued on next page]

(54) Title: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE

(57) Abstract: Disclosed herein are nucleic acid sequences that encode novel polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies that immunospecifically bind to the polypeptide, as well as derivatives, variants, mutants, or fragments of the novel polypeptide, polynucleotide, or antibody specific to the polypeptide. Vectors, host cells, antibodies and recombinant methods for producing the polypeptides and polynucleotides, as well as methods for using same are also included. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.



06437 (US). **SHENOY, Suresh, G.** [IN/US]; 15 Millwood Drive, Branford, CT 06405 (US). **TAUPIER, Raymond, J., JR.** [US/US]; 34 Pardee Place Extension, East Haven, CT 06512 (US). **VERNET, Corine, A.M.** [FR/US]; 1739 Foxon Road, Apartment L6, Branford, CT 06471 (US). **ZHONG, Mei** [CA/US]; 45 Harrison Avenue, Apartment 1B, Branford, CT 06405 (US). **MALYANKAR, Uriel, M.** [IN/US]; 229 Branford Road, Number 330, Branford, CT 06405 (US). **MILLET, Isabelle** [FR/US]; 74 Carrington Avenue, Milford, CT 06460 (US). **KEKUDA, Ramesh** [IN/US]; 71 Aiken Street, Unit R3, Norwalk, CT 06851 (US).

(74) **Agent: ELRIFI, Ivor, R. Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C.**; One Financial Center, Boston, MA 02111 (US).

(81) **Designated States (national):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,

LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NI, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## **THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE**

### **FIELD OF THE INVENTION**

The present invention relates to novel polypeptides, and the nucleic acids encoding them, having properties related to stimulation of biochemical or physiological responses in a cell, a tissue, an organ or an organism. More particularly, the novel polypeptides are gene products of novel genes, or are specified biologically active fragments or derivatives thereof. Methods of use encompass diagnostic and prognostic assay procedures as well as methods of treating diverse pathological conditions.

## BACKGROUND OF THE INVENTION

Eukaryotic cells are characterized by biochemical and physiological processes which under normal conditions are exquisitely balanced to achieve the preservation and propagation of the cells. When such cells are components of multicellular organisms such as vertebrates, or more particularly organisms such as mammals, the regulation of the biochemical and physiological processes involves intricate signaling pathways. Frequently, such signaling pathways involve extracellular signaling proteins, cellular receptors that bind the signaling proteins, and signal transducing components located within the cells.

Signaling proteins may be classified as endocrine effectors, paracrine effectors or autocrine effectors. Endocrine effectors are signaling molecules secreted by a given organ into the circulatory system, which are then transported to a distant target organ or tissue. The target cells include the receptors for the endocrine effector, and when the endocrine effector binds, a signaling cascade is induced. Paracrine effectors involve secreting cells and receptor cells in close proximity to each other, for example two different classes of cells in the same tissue or organ. One class of cells secretes the paracrine effector, which then reaches the second class of cells, for example by diffusion through the extracellular fluid. The second class of cells contains the receptors for the paracrine effector; binding of the effector results in induction of the signaling cascade that elicits the corresponding biochemical or physiological effect. Autocrine effectors are highly analogous to paracrine effectors, except that the same cell type that secretes the autocrine effector also contains the receptor. Thus the autocrine effector binds to receptors on the same cell, or on identical neighboring cells. The binding process then elicits the characteristic biochemical or physiological effect.

Signaling processes may elicit a variety of effects on cells and tissues including by way of nonlimiting example induction of cell or tissue proliferation, suppression of growth or proliferation, induction of differentiation or maturation of a cell or tissue, and suppression of differentiation or maturation of a cell or tissue.

Many pathological conditions involve dysregulation of expression of important effector proteins. In certain classes of pathologies the dysregulation is manifested as diminished or suppressed level of synthesis and secretion of protein effectors. In other classes of pathologies the dysregulation is manifested as increased or up-regulated level of synthesis and secretion of protein effectors. In a clinical setting a subject may be suspected

of suffering from a condition brought on by altered or mis-regulated levels of a protein effector of interest. Therefore there is a need to assay for the level of the protein effector of interest in a biological sample from such a subject, and to compare the level with that characteristic of a nonpathological condition. There also is a need to provide the protein effector as a product of manufacture. Administration of the effector to a subject in need thereof is useful in treatment of the pathological condition. Accordingly, there is a need for a method of treatment of a pathological condition brought on by a diminished or suppressed levels of the protein effector of interest. In addition, there is a need for a method of treatment of a pathological condition brought on by a increased or up-regulated levels of the protein effector of interest.

Antibodies are multichain proteins that bind specifically to a given antigen, and bind poorly, or not at all, to substances deemed not to be cognate antigens. Antibodies are comprised of two short chains termed light chains and two long chains termed heavy chains. These chains are constituted of immunoglobulin domains, of which generally there are two classes: one variable domain per chain, one constant domain in light chains, and three or more constant domains in heavy chains. The antigen-specific portion of the immunoglobulin molecules resides in the variable domains; the variable domains of one light chain and one heavy chain associate with each other to generate the antigen-binding moiety. Antibodies that bind immunospecifically to a cognate or target antigen bind with high affinities. Accordingly, they are useful in assaying specifically for the presence of the antigen in a sample. In addition, they have the potential of inactivating the activity of the antigen.

Therefore there is a need to assay for the level of a protein effector of interest in a biological sample from such a subject, and to compare this level with that characteristic of a nonpathological condition. In particular, there is a need for such an assay based on the use of an antibody that binds immunospecifically to the antigen. There further is a need to inhibit the activity of the protein effector in cases where a pathological condition arises from elevated or excessive levels of the effector based on the use of an antibody that binds immunospecifically to the effector. Thus, there is a need for the antibody as a product of manufacture. There further is a need for a method of treatment of a pathological condition brought on by an elevated or excessive level of the protein effector of interest based on administering the antibody to the subject.

## SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of isolated polypeptides including amino acid sequences selected from mature forms of the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52. The novel nucleic acids and polypeptides are referred to herein as NOVX, where X is an identifier for each sequence as shown in Table A below. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "NOVX" nucleic acid or polypeptide sequences.

The invention also is based in part upon variants of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed. In another embodiment, the invention includes the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52. In another embodiment, the invention also comprises variants of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52, wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed. The invention also involves fragments of any of the mature forms of the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52, or any other amino acid sequence selected from this group. The invention also comprises fragments from these groups in which up to 15% of the residues are changed.

In another embodiment, the invention encompasses polypeptides that are naturally occurring allelic variants of the sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52. These allelic variants include amino acid sequences that are the translations of nucleic acid sequences differing by a single nucleotide from nucleic acid sequences selected from the group consisting of SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 52. The variant polypeptide where any amino acid changed in the chosen sequence is changed to provide a conservative substitution.

In another embodiment, the invention comprises a pharmaceutical composition involving a polypeptide with an amino acid sequence selected from the group consisting of

SEQ ID NO:2n, wherein n is an integer between 1 and 52 and a pharmaceutically acceptable carrier. In another embodiment, the invention involves a kit, including, in one or more containers, this pharmaceutical composition.

In another embodiment, the invention includes the use of a therapeutic in the  
5 manufacture of a medicament for treating a syndrome associated with a human disease, the disease being selected from a pathology associated with a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52 wherein said therapeutic is the polypeptide selected from this group. In another embodiment, the invention comprises a method for determining the presence or  
10 amount of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52 in a sample, the method involving providing the sample; introducing the sample to an antibody that binds immunospecifically to the polypeptide; and determining the presence or amount of antibody bound to the polypeptide, thereby determining the presence or amount of  
15 polypeptide in the sample.

In another embodiment, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52 in a first mammalian subject, the method  
20 involving measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and comparing the amount of the polypeptide in this sample to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, the disease, wherein an alteration in the expression level of the polypeptide in the first subject as compared to the control sample  
25 indicates the presence of or predisposition to the disease.

In another embodiment, the invention involves a method of identifying an agent that binds to a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52, the method including  
30 introducing the polypeptide to the agent; and determining whether the agent binds to the polypeptide. The agent could be a cellular receptor or a downstream effector.

In another embodiment, the invention involves a method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer

between 1 and 52, the method including providing a cell expressing the polypeptide of the invention and having a property or function ascribable to the polypeptide; contacting the cell with a composition comprising a candidate substance; and determining whether the substance alters the property or function ascribable to the polypeptide; whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic agent.

In another embodiment, the invention involves a method for screening for a modulator of activity or of latency or predisposition to a pathology associated with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52, the method including administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of the invention, wherein the test animal recombinantly expresses the polypeptide of the invention; measuring the activity of the polypeptide in the test animal after administering the test compound; and comparing the activity of the protein in the test animal with the activity of the polypeptide in a control animal not administered the polypeptide, wherein a change in the activity of the polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of, or predisposition to, a pathology associated with the polypeptide of the invention. The recombinant test animal could express a test protein transgene or express the transgene under the control of a promoter at an increased level relative to a wild-type test animal. The promoter may or may not be the native gene promoter of the transgene.

In another embodiment, the invention involves a method for modulating the activity of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52, the method including introducing a cell sample expressing the polypeptide with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide.

In another embodiment, the invention involves a method of treating or preventing a pathology associated with a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52, the method including administering the polypeptide to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject. The subject could be human.

In another embodiment, the invention involves a method of treating a pathological state in a mammal, the method including administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52 or a biologically active fragment thereof.

In another embodiment, the invention involves an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide having an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 52; a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52; a variant of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52 or any variant of the polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and the complement of any of the nucleic acid molecules.

In another embodiment, the invention comprises an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 52, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant.

In another embodiment, the invention involves an isolated nucleic acid molecule including a nucleic acid sequence encoding a polypeptide having an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 52 that encodes a variant polypeptide,

wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant.

In another embodiment, the invention comprises an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence  
5 selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 52, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 52.

In another embodiment, the invention includes an isolated nucleic acid molecule  
10 having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 52, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer  
15 between 1 and 52; a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; a nucleic acid fragment of the sequence selected from the group consisting of  
20 SEQ ID NO:2n-1, wherein n is an integer between 1 and 52; and a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

25 In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 52, wherein the nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group  
30 consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, or a complement of the nucleotide sequence.

In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ

ID NO:2n, wherein n is an integer between 1 and 52, wherein the nucleic acid molecule has a nucleotide sequence in which any nucleotide specified in the coding sequence of the chosen nucleotide sequence is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the  
5 nucleotides in the chosen coding sequence are so changed, an isolated second polynucleotide that is a complement of the first polynucleotide, or a fragment of any of them.

In another embodiment, the invention includes a vector involving the nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid  
10 sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 52. This vector can have a promoter operably linked to the nucleic acid molecule. This vector can be located within a cell.

In another embodiment, the invention involves a method for determining the  
15 presence or amount of a nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 52 in a sample, the method including providing the sample; introducing the sample to a probe that binds to the nucleic acid molecule; and determining the presence or  
20 amount of the probe bound to the nucleic acid molecule, thereby determining the presence or amount of the nucleic acid molecule in the sample. The presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type. The cell type can be cancerous.

In another embodiment, the invention involves a method for determining the  
25 presence of or predisposition for a disease associated with altered levels of a nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 52 in a first mammalian  
subject, the method including measuring the amount of the nucleic acid in a sample from  
30 the first mammalian subject; and comparing the amount of the nucleic acid in the sample of step (a) to the amount of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease; wherein an alteration in the level of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

The invention further provides an antibody that binds immunospecifically to a NOVX polypeptide. The NOVX antibody may be monoclonal, humanized, or a fully human antibody. Preferably, the antibody has a dissociation constant for the binding of the NOVX polypeptide to the antibody less than  $1 \times 10^{-9}$  M. More preferably, the NOVX antibody neutralizes the activity of the NOVX polypeptide.

In a further aspect, the invention provides for the use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, associated with a NOVX polypeptide. Preferably the therapeutic is a NOVX antibody. In yet a further aspect, the invention provides a method of treating or preventing a NOVX-associated disorder, a method of treating a pathological state in a mammal, and a method of treating or preventing a pathology associated with a polypeptide by administering a NOVX antibody to a subject in an amount sufficient to treat or prevent the disorder.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

25

## DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences, their encoded polypeptides, antibodies, and other related compounds. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel

sequences disclosed herein. Table A provides a summary of the NOVX nucleic acids and their encoded polypeptides.

**TABLE A. SEQUENCES AND CORRESPONDING SEQ ID NUMBERS**

5

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (amino acid)	Homology
NOV1a	CG108030-01	1	2	Human Sequence
NOV1b	CG108030-02	3	4	Human Sequence
NOV2a	CG115907-01	5	6	Trypsin inhibitor precursor
NOV2b	CG115907-04	7	8	Trypsin inhibitor precursor
NOV2c	CG115907-03	9	10	Trypsin inhibitor precursor
NOV2d	CG115907-02	11	12	Trypsin inhibitor precursor
NOV3a	CG139008-01	13	14	Binding protein
NOV3b	233028732	15	16	Binding protein
NOV3c	CG139008-02	17	18	Binding protein
NOV4a	CG145877-01	19	20	Hypothetical protein
NOV5a	CG151161-02	21	22	Myelin and lymphocyte protein
NOV5b	CG151161-01	23	24	Myelin and lymphocyte protein
NOV6a	CG155653-01	25	26	Similar to PDZ domain
NOV7a	CG160093-01	27	28	Leukocyte elastase inhibitor
NOV7b	CG160093-02	29	30	Leukocyte elastase inhibitor
NOV8a	CG163133-02	31	32	JM4 protein
NOV8b	CG163133-01	33	34	JM4 protein
NOV9a	CG165528-01	35	36	Neurexin 1-alpha precursor
NOV9b	CG165528-02	37	38	Neurexin 1-alpha precursor
NOV10a	CG165666-01	39	40	Similar to TPR-containing protein
NOV11a	CG165676-01	41	42	Integrin, alpha 2
NOV12a	CG165719-04	43	44	Neuronal membrane protein M6-B
NOV12b	CG165719-02	45	46	Neuronal membrane protein M6-B
NOV12c	CG165719-03	47	48	Neuronal membrane protein M6-B
NOV12d	CG165719-01	49	50	Neuronal membrane protein M6-B
NOV12e	CG165719-05	51	52	Neuronal membrane protein M6-B
NOV13a	CG167488-02	53	54	Human protein
NOV13b	CG167488-01	55	56	Human protein
NOV14a	CG173318-01	57	58	Human protein
NOV15a	CG50970-06	59	60	cerebroglycan
NOV15b	CG50970-01	61	62	cerebroglycan
NOV15d	274054257	63	64	cerebroglycan
NOV15e	CG50970-03	65	66	cerebroglycan
NOV15f	237922026	67	68	cerebroglycan
NOV15g	237922511	69	70	cerebroglycan
NOV15h	315490136	71	72	cerebroglycan

NOV15i	CG50970-02	73	74	cerebroglycan
NOV15j	CG50970-04	75	76	cerebroglycan
NOV15k	CG50970-05	77	78	cerebroglycan
NOV15l	CG50970-07	79	80	cerebroglycan
NOV16a	CG54443-03	81	82	Hypothetical Protein
NOV16b	CG54443-07	83	84	Hypothetical Protein
NOV16c	CG54443-01	85	86	Hypothetical Protein
NOV16d	CG54443-02	87	88	Hypothetical Protein
NOV16e	CG54443-04	89	90	Hypothetical Protein
NOV16f	CG54443-05	91	92	Hypothetical Protein
NOV16g	CG54443-06	93	94	Hypothetical Protein
NOV17a	CG58495-01	95	96	pulmonary surfactant protein
NOV17b	CG58495-03	97	98	pulmonary surfactant protein
NOV17c	CG58495-02	99	100	pulmonary surfactant protein
NOV18a	CG97482-01	101	102	S-100 protein, beta chain
NOV18b	CG97482-02	103	104	S-100 protein, beta chain

Table A indicates the homology of NOVX polypeptides to known protein families. Thus, the nucleic acids and polypeptides, antibodies and related compounds according to the invention corresponding to a NOVX as identified in column 1 of Table A will be useful in therapeutic and diagnostic applications implicated in, for example, pathologies and disorders associated with the known protein families identified in column 5 of Table A.

Pathologies, diseases, disorders and conditions and the like that are associated with NOVX sequences include, but are not limited to: *e.g.*, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), vascular calcification, fibrosis, atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, metabolic disturbances associated with obesity, transplantation, osteoarthritis, rheumatoid arthritis, osteochondrodysplasia, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, glomerulonephritis, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, psoriasis, skin disorders, graft versus host disease, AIDS, bronchial asthma, lupus, Crohn's disease; inflammatory bowel disease, ulcerative colitis, multiple sclerosis, treatment of Albright Hereditary Osteodystrophy, infectious disease, anorexia, cancer-associated cachexia; cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias, schizophrenia, depression, asthma, emphysema, allergies, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers, as

well as conditions such as transplantation, neuroprotection, fertility, or regeneration (*in vitro* and *in vivo*).

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

Consistent with other known members of the family of proteins, identified in column 5 of Table A, the NOVX polypeptides of the present invention show homology to, and contain domains that are characteristic of, other members of such protein families. Details of the sequence relatedness and domain analysis for each NOVX are presented in Example A.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit diseases associated with the protein families listed in Table A.

The NOVX nucleic acids and polypeptides are also useful for detecting specific cell types. Details of the expression analysis for each NOVX are presented in Example C. Accordingly, the NOVX nucleic acids, polypeptides, antibodies and related compounds according to the invention will have diagnostic and therapeutic applications in the detection of a variety of diseases with differential expression in normal vs. diseased tissues, *e.g.* detection of a variety of cancers.

Additional utilities for NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

#### **NOVX clones**

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

The NOVX genes and their corresponding encoded proteins are useful for preventing, treating or ameliorating medical conditions, *e.g.*, by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are  
5 described for each of the NOVX genes, based on the tissues in which they are most highly expressed. Uses include developing products for the diagnosis or treatment of a variety of diseases and disorders.

The NOVX nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a  
10 specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene  
15 ablation), and (v) a composition promoting tissue regeneration *in vitro* and *in vivo* (vi) a biological defense weapon.

In one specific embodiment, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n,  
20 wherein n is an integer between 1 and 52; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 52, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) an amino acid sequence selected from the group  
25 consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 52; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 52, wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and (e) a fragment of any of (a) through (d).

30 In another specific embodiment, the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 52; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID

NO: 2n, wherein n is an integer between 1 and 52, wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 52; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 52, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; (e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 52, or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and (f) the complement of any of said nucleic acid molecules.

In yet another specific embodiment, the invention includes an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 52; (b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 52, is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; (c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 52; and (d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 52, is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

#### **NOVX Nucleic Acids and Polypeptides**

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR

primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The  
5 nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

A NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally  
10 occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product "mature" form arises, by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell (*e.g.*, host  
15 cell) in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal  
20 methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of  
25 post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probe", as utilized herein, refers to nucleic acid sequences of variable  
30 length, preferably between at least about 10 nucleotides (nt), about 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single-

stranded or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as used herein, is a nucleic acid that is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, *etc.*). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium, or of chemical precursors or other chemicals.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 52, or a complement of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 52, as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template with appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer. As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt

to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of a NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, is one that is sufficiently complementary to the nucleotide sequence of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, that it can hydrogen bond with few or no mismatches to the nucleotide sequence shown in SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

A "fragment" provided herein is defined as a sequence of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, and is at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice.

A full-length NOVX clone is identified as containing an ATG translation start codon and an in-frame stop codon. Any disclosed NOVX nucleotide sequence lacking an ATG start codon therefore encodes a truncated C-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 5'

direction of the disclosed sequence. Any disclosed NOVX nucleotide sequence lacking an in-frame stop codon similarly encodes a truncated N-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 3' direction of the disclosed sequence.

5           A "derivative" is a nucleic acid sequence or amino acid sequence formed from the native compounds either directly, by modification or partial substitution. An "analog" is a nucleic acid sequence or amino acid sequence that has a structure similar to, but not identical to, the native compound, *e.g.* they differs from it in respect to certain components or side chains. Analogs may be synthetic or derived from a different evolutionary origin  
10          and may have a similar or opposite metabolic activity compared to wild type. A "homolog" is a nucleic acid sequence or amino acid sequence of a particular gene that is derived from different species.

Derivatives and analogs may be full length or other than full length. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to,  
15          molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable  
20          of hybridizing to the complement of a sequence encoding the proteins under stringent, moderately stringent, or low stringent conditions. *See e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or  
25          variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences include those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention,  
30          homologous nucleotide sequences include nucleotide sequences encoding for a NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous

nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, as well as a polypeptide possessing  
 5 NOVX biological activity. Various biological activities of the NOVX proteins are described below.

A NOVX polypeptide is encoded by the open reading frame ("ORF") of a NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is  
 10 uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a  
 15 minimum size requirement is often set, *e.g.*, a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX  
 20 homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52; or an anti-sense strand nucleotide  
 25 sequence of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52; or of a naturally occurring mutant of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52. Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe has a detectable label attached, *e.g.* the label can be a radioisotope, a fluorescent  
 30 compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express a NOVX protein, such as by measuring a level of a NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of a NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a

- 5 "biologically-active portion of NOVX" can be prepared by isolating a portion of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, that encodes a polypeptide having a NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

10 **NOVX Nucleic Acid and Polypeptide Variants**

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52.

- 15 In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence of SEQ ID NO:2n, wherein n is an integer between 1 and 52.

In addition to the human NOVX nucleotide sequences of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding a NOVX

20 protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

- 25
- 30 Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from a human SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the

NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

5           Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 52. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another  
10       embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 65% homologous to each other typically remain hybridized to each other.

          Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other  
15       than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning. As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no  
20       other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (*T<sub>m</sub>*) for the specific sequence at a defined ionic strength and pH. The *T<sub>m</sub>* is the temperature (under defined ionic strength, pH and nucleic acid  
25       concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at *T<sub>m</sub>*, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the  
30       temperature is at least about 30 °C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60 °C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Reinhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55 °C, followed by one or more washes in 1X SSC, 0.1% SDS at 37 °C. Other conditions of moderate stringency that may be used are well-known within the art. *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Krieger, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND

EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981.  
*Proc Natl Acad Sci USA* 78: 6789-6792.

### Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may  
5 exist in the population, the skilled artisan will further appreciate that changes can be  
introduced by mutation into the nucleotide sequences of SEQ ID NO:2n-1, wherein n is an  
integer between 1 and 52, thereby leading to changes in the amino acid sequences of the  
encoded NOVX protein, without altering the functional ability of that NOVX protein. For  
example, nucleotide substitutions leading to amino acid substitutions at "non-essential"  
10 amino acid residues can be made in the sequence of SEQ ID NO:2n, wherein n is an integer  
between 1 and 52. A "non-essential" amino acid residue is a residue that can be altered  
from the wild-type sequences of the NOVX proteins without altering their biological  
activity, whereas an "essential" amino acid residue is required for such biological activity.  
For example, amino acid residues that are conserved among the NOVX proteins of the  
15 invention are predicted to be particularly non-amenable to alteration. Amino acids for  
which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX  
proteins that contain changes in amino acid residues that are not essential for activity. Such  
NOVX proteins differ in amino acid sequence from SEQ ID NO:2n-1, wherein n is an  
20 integer between 1 and 52, yet retain biological activity. In one embodiment, the isolated  
nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the  
protein comprises an amino acid sequence at least about 40% homologous to the amino  
acid sequences of SEQ ID NO:2n, wherein n is an integer between 1 and 52. Preferably,  
the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ  
25 ID NO:2n, wherein n is an integer between 1 and 52; more preferably at least about 70%  
homologous to SEQ ID NO:2n, wherein n is an integer between 1 and 52; still more  
preferably at least about 80% homologous to SEQ ID NO:2n, wherein n is an integer  
between 1 and 52; even more preferably at least about 90% homologous to SEQ ID NO:2n,  
wherein n is an integer between 1 and 52; and most preferably at least about 95%  
30 homologous to SEQ ID NO:2n, wherein n is an integer between 1 and 52.

An isolated nucleic acid molecule encoding a NOVX protein homologous to the  
protein of SEQ ID NO:2n, wherein n is an integer between 1 and 52, can be created by  
introducing one or more nucleotide substitutions, additions or deletions into the nucleotide

sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 52, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

5 Mutations can be introduced any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 52, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains  
10 have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine).  
15 Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for  
20 NOVX biological activity to identify mutants that retain activity. Following mutagenesis of a nucleic acid of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 52, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain  
25 interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may  
30 be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant NOVX protein can be assayed for (*i*) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or

biologically-active portions thereof, (ii) complex formation between a mutant NOVX protein and a NOVX ligand; or (iii) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (e.g. avidin proteins). In yet another embodiment, a mutant NOVX protein can be assayed for the ability to  
5 regulate a specific biological function (e.g., regulation of insulin release).

### Interfering RNA

In one aspect of the invention, NOVX gene expression can be attenuated by RNA interference. One approach well-known in the art is short interfering RNA (siRNA) mediated gene silencing where expression products of a NOVX gene are targeted by  
10 specific double stranded NOVX derived siRNA nucleotide sequences that are complementary to at least a 19-25 nt long segment of the NOVX gene transcript, including the 5' untranslated (UT) region, the ORF, or the 3' UT region. See, e.g., PCT applications WO00/44895, WO99/32619, WO01/75164, WO01/92513, WO 01/29058, WO01/89304, WO02/16620, and WO02/29858, each incorporated by reference herein in their entirety.  
15 Targeted genes can be a NOVX gene, or an upstream or downstream modulator of the NOVX gene. Nonlimiting examples of upstream or downstream modulators of a NOVX gene include, e.g., a transcription factor that binds the NOVX gene promoter, a kinase or phosphatase that interacts with a NOVX polypeptide, and polypeptides involved in a NOVX regulatory pathway.

20 According to the methods of the present invention, NOVX gene expression is silenced using short interfering RNA. A NOVX polynucleotide according to the invention includes a siRNA polynucleotide. Such a NOVX siRNA can be obtained using a NOVX polynucleotide sequence, for example, by processing the NOVX ribopolynucleotide sequence in a cell-free system, such as but not limited to a Drosophila extract, or by  
25 transcription of recombinant double stranded NOVX RNA or by chemical synthesis of nucleotide sequences homologous to a NOVX sequence. See, e.g., Tuschl, Zamore, Lehmann, Bartel and Sharp (1999), Genes & Dev. 13: 3191-3197, incorporated herein by reference in its entirety. When synthesized, a typical 0.2 micromolar-scale RNA synthesis provides about 1 milligram of siRNA, which is sufficient for 1000 transfection experiments  
30 using a 24-well tissue culture plate format.

The most efficient silencing is generally observed with siRNA duplexes composed of a 21-nt sense strand and a 21-nt antisense strand, paired in a manner to have a 2-nt 3' overhang. The sequence of the 2-nt 3' overhang makes an additional small contribution

to the specificity of siRNA target recognition. The contribution to specificity is localized to the unpaired nucleotide adjacent to the first paired bases. In one embodiment, the nucleotides in the 3' overhang are ribonucleotides. In an alternative embodiment, the nucleotides in the 3' overhang are deoxyribonucleotides. Using 2'-deoxyribonucleotides in the 3' overhangs is as efficient as using ribonucleotides, but deoxyribonucleotides are often cheaper to synthesize and are most likely more nuclease resistant.

A contemplated recombinant expression vector of the invention comprises a NOVX DNA molecule cloned into an expression vector comprising operatively-linked regulatory sequences flanking the NOVX sequence in a manner that allows for expression (by transcription of the DNA molecule) of both strands. An RNA molecule that is antisense to NOVX mRNA is transcribed by a first promoter (*e.g.*, a promoter sequence 3' of the cloned DNA) and an RNA molecule that is the sense strand for the NOVX mRNA is transcribed by a second promoter (*e.g.*, a promoter sequence 5' of the cloned DNA). The sense and antisense strands may hybridize *in vivo* to generate siRNA constructs for silencing of the NOVX gene. Alternatively, two constructs can be utilized to create the sense and anti-sense strands of a siRNA construct. Finally, cloned DNA can encode a construct having secondary structure, wherein a single transcript has both the sense and complementary antisense sequences from the target gene or genes. In an example of this embodiment, a hairpin RNAi product is homologous to all or a portion of the target gene. In another example, a hairpin RNAi product is a siRNA. The regulatory sequences flanking the NOVX sequence may be identical or may be different, such that their expression may be modulated independently, or in a temporal or spatial manner.

In a specific embodiment, siRNAs are transcribed intracellularly by cloning the NOVX gene templates into a vector containing, *e.g.*, a RNA pol III transcription unit from the smaller nuclear RNA (snRNA) U6 or the human RNase P RNA H1. One example of a vector system is the GeneSuppressor<sup>TM</sup> RNA Interference kit (commercially available from Imgenex). The U6 and H1 promoters are members of the type III class of Pol III promoters. The +1 nucleotide of the U6-like promoters is always guanosine, whereas the +1 for H1 promoters is adenosine. The termination signal for these promoters is defined by five consecutive thymidines. The transcript is typically cleaved after the second uridine. Cleavage at this position generates a 3' UU overhang in the expressed siRNA, which is similar to the 3' overhangs of synthetic siRNAs. Any sequence less than 400 nucleotides in length can be transcribed by these promoter, therefore they are ideally suited for the

expression of around 21-nucleotide siRNAs in, *e.g.*, an approximately 50-nucleotide RNA stem-loop transcript.

A siRNA vector appears to have an advantage over synthetic siRNAs where long term knock-down of expression is desired. Cells transfected with a siRNA expression  
5 vector would experience steady, long-term mRNA inhibition. In contrast, cells transfected with exogenous synthetic siRNAs typically recover from mRNA suppression within seven days or ten rounds of cell division. The long-term gene silencing ability of siRNA expression vectors may provide for applications in gene therapy.

In general, siRNAs are chopped from longer dsRNA by an ATP-dependent  
10 ribonuclease called DICER. DICER is a member of the RNase III family of double-stranded RNA-specific endonucleases. The siRNAs assemble with cellular proteins into an endonuclease complex. *In vitro* studies in *Drosophila* suggest that the siRNAs/protein complex (siRNP) is then transferred to a second enzyme complex, called an RNA-induced silencing complex (RISC), which contains an endoribonuclease that is  
15 distinct from DICER. RISC uses the sequence encoded by the antisense siRNA strand to find and destroy mRNAs of complementary sequence. The siRNA thus acts as a guide, restricting the ribonuclease to cleave only mRNAs complementary to one of the two siRNA strands.

A NOVX mRNA region to be targeted by siRNA is generally selected from a  
20 desired NOVX sequence beginning 50 to 100 nt downstream of the start codon. Alternatively, 5' or 3' UTRs and regions nearby the start codon can be used but are generally avoided, as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. An initial BLAST homology search for the selected siRNA  
25 sequence is done against an available nucleotide sequence library to ensure that only one gene is targeted. Specificity of target recognition by siRNA duplexes indicate that a single point mutation located in the paired region of an siRNA duplex is sufficient to abolish target mRNA degradation. See, Elbashir *et al.* 2001 EMBO J. 20(23):6877-88. Hence, consideration should be taken to accommodate SNPs, polymorphisms, allelic variants or  
30 species-specific variations when targeting a desired gene.

In one embodiment, a complete NOVX siRNA experiment includes the proper negative control. A negative control siRNA generally has the same nucleotide composition as the NOVX siRNA but lack significant sequence homology to the genome. Typically,

one would scramble the nucleotide sequence of the NOVX siRNA and do a homology search to make sure it lacks homology to any other gene.

Two independent NOVX siRNA duplexes can be used to knock-down a target NOVX gene. This helps to control for specificity of the silencing effect. In addition,  
5 expression of two independent genes can be simultaneously knocked down by using equal concentrations of different NOVX siRNA duplexes, *e.g.*, a NOVX siRNA and an siRNA for a regulator of a NOVX gene or polypeptide. Availability of siRNA-associating proteins is believed to be more limiting than target mRNA accessibility.

A targeted NOVX region is typically a sequence of two adenines (AA) and two  
10 thymidines (TT) divided by a spacer region of nineteen (N19) residues (*e.g.*, AA(N19)TT). A desirable spacer region has a G/C-content of approximately 30% to 70%, and more preferably of about 50%. If the sequence AA(N19)TT is not present in the target sequence, an alternative target region would be AA(N21). The sequence of the NOVX sense siRNA corresponds to (N19)TT or N21, respectively. In the latter case, conversion of the 3' end of  
15 the sense siRNA to TT can be performed if such a sequence does not naturally occur in the NOVX polynucleotide. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and antisense 3' overhangs. Symmetric 3' overhangs may help to ensure that the siRNPs are formed with approximately equal ratios of sense and antisense target RNA-cleaving siRNPs. *See, e.g.*,  
20 Elbashir, Lendeckel and Tuschl (2001). *Genes & Dev.* 15: 188-200, incorporated by reference herein in its entirety. The modification of the overhang of the sense sequence of the siRNA duplex is not expected to affect targeted mRNA recognition, as the antisense siRNA strand guides target recognition.

Alternatively, if the NOVX target mRNA does not contain a suitable AA(N21)  
25 sequence, one may search for the sequence NA(N21). Further, the sequence of the sense strand and antisense strand may still be synthesized as 5' (N19)TT, as it is believed that the sequence of the 3'-most nucleotide of the antisense siRNA does not contribute to specificity. Unlike antisense or ribozyme technology, the secondary structure of the target mRNA does not appear to have a strong effect on silencing. *See, Harborth, et al.* (2001) *J. Cell Science* 114: 4557-4565, incorporated by reference in its entirety.  
30

Transfection of NOVX siRNA duplexes can be achieved using standard nucleic acid transfection methods, for example, OLIGOFECTAMINE Reagent (commercially available from Invitrogen). An assay for NOVX gene silencing is generally performed approximately 2 days after transfection. No NOVX gene silencing has been observed in

the absence of transfection reagent, allowing for a comparative analysis of the wild-type and silenced NOVX phenotypes. In a specific embodiment, for one well of a 24-well plate, approximately 0.84  $\mu$ g of the siRNA duplex is generally sufficient. Cells are typically seeded the previous day, and are transfected at about 50% confluence. The choice of cell culture media and conditions are routine to those of skill in the art, and will vary with the choice of cell type. The efficiency of transfection may depend on the cell type, but also on the passage number and the confluency of the cells. The time and the manner of formation of siRNA-liposome complexes (*e.g.* inversion versus vortexing) are also critical. Low transfection efficiencies are the most frequent cause of unsuccessful NOVX silencing. The efficiency of transfection needs to be carefully examined for each new cell line to be used. Preferred cell are derived from a mammal, more preferably from a rodent such as a rat or mouse, and most preferably from a human. Where used for therapeutic treatment, the cells are preferentially autologous, although non-autologous cell sources are also contemplated as within the scope of the present invention.

For a control experiment, transfection of 0.84  $\mu$ g single-stranded sense NOVX siRNA will have no effect on NOVX silencing, and 0.84  $\mu$ g antisense siRNA has a weak silencing effect when compared to 0.84  $\mu$ g of duplex siRNAs. Control experiments again allow for a comparative analysis of the wild-type and silenced NOVX phenotypes. To control for transfection efficiency, targeting of common proteins is typically performed, for example targeting of lamin A/C or transfection of a CMV-driven EGFP-expression plasmid (*e.g.* commercially available from Clontech). In the above example, a determination of the fraction of lamin A/C knockdown in cells is determined the next day by such techniques as immunofluorescence, Western blot, Northern blot or other similar assays for protein expression or gene expression. Lamin A/C monoclonal antibodies may be obtained from Santa Cruz Biotechnology.

Depending on the abundance and the half life (or turnover) of the targeted NOVX polynucleotide in a cell, a knock-down phenotype may become apparent after 1 to 3 days, or even later. In cases where no NOVX knock-down phenotype is observed, depletion of the NOVX polynucleotide may be observed by immunofluorescence or Western blotting. If the NOVX polynucleotide is still abundant after 3 days, cells need to be split and transferred to a fresh 24-well plate for re-transfection. If no knock-down of the targeted protein is observed, it may be desirable to analyze whether the target mRNA (NOVX or a NOVX upstream or downstream gene) was effectively destroyed by the transfected siRNA duplex. Two days after transfection, total RNA is prepared, reverse transcribed using a

target-specific primer, and PCR-amplified with a primer pair covering at least one exon-exon junction in order to control for amplification of pre-mRNAs. RT/PCR of a non-targeted mRNA is also needed as control. Effective depletion of the mRNA yet undetectable reduction of target protein may indicate that a large reservoir of stable NOVX  
5 protein may exist in the cell. Multiple transfection in sufficiently long intervals may be necessary until the target protein is finally depleted to a point where a phenotype may become apparent. If multiple transfection steps are required, cells are split 2 to 3 days after transfection. The cells may be transfected immediately after splitting.

An inventive therapeutic method of the invention contemplates administering a  
10 NOVX siRNA construct as therapy to compensate for increased or aberrant NOVX expression or activity. The NOVX ribopolynucleotide is obtained and processed into siRNA fragments, or a NOVX siRNA is synthesized, as described above. The NOVX siRNA is administered to cells or tissues using known nucleic acid transfection techniques, as described above. A NOVX siRNA specific for a NOVX gene will decrease or  
15 knockdown NOVX transcription products, which will lead to reduced NOVX polypeptide production, resulting in reduced NOVX polypeptide activity in the cells or tissues. The present invention also encompasses a method of treating a disease or condition associated with the presence of a NOVX protein in an individual comprising administering to the individual an RNAi construct that targets the mRNA of the protein (the mRNA that  
20 encodes the protein) for degradation. A specific RNAi construct includes a siRNA or a double stranded gene transcript that is processed into siRNAs. Upon treatment, the target protein is not produced or is not produced to the extent it would be in the absence of the treatment.

Where the NOVX gene function is not correlated with a known phenotype, a  
25 control sample of cells or tissues from healthy individuals provides a reference standard for determining NOVX expression levels. Expression levels are detected using the assays described, *e.g.*, RT-PCR, Northern blotting, Western blotting, ELISA, and the like. A subject sample of cells or tissues is taken from a mammal, preferably a human subject, suffering from a disease state. The NOVX ribopolynucleotide is used to produce siRNA  
30 constructs, that are specific for the NOVX gene product. These cells or tissues are treated by administering NOVX siRNA's to the cells or tissues by methods described for the transfection of nucleic acids into a cell or tissue, and a change in NOVX polypeptide or polynucleotide expression is observed in the subject sample relative to the control sample, using the assays described. This NOVX gene knockdown approach provides a rapid

method for determination of a NOVX minus (NOVX<sup>-</sup>) phenotype in the treated subject sample. The NOVX<sup>-</sup> phenotype observed in the treated subject sample thus serves as a marker for monitoring the course of a disease state during treatment.

In specific embodiments, a NOVX siRNA is used in therapy. Methods for the generation and use of a NOVX siRNA are known to those skilled in the art. Example techniques are provided below.

### Production of RNAs

Sense RNA (ssRNA) and antisense RNA (asRNA) of NOVX are produced using known methods such as transcription in RNA expression vectors. In the initial experiments, the sense and antisense RNA are about 500 bases in length each. The produced ssRNA and asRNA (0.5  $\mu$ M) in 10 mM Tris-HCl (pH 7.5) with 20 mM NaCl were heated to 95° C for 1 min then cooled and annealed at room temperature for 12 to 16 h. The RNAs are precipitated and resuspended in lysis buffer (below). To monitor annealing, RNAs are electrophoresed in a 2% agarose gel in TBE buffer and stained with ethidium bromide. See, *e.g.*, Sambrook et al., Molecular Cloning. Cold Spring Harbor Laboratory Press, Plainview, N.Y. (1989).

### Lysate Preparation

Untreated rabbit reticulocyte lysate (Ambion) are assembled according to the manufacturer's directions. dsRNA is incubated in the lysate at 30° C for 10 min prior to the addition of mRNAs. Then NOVX mRNAs are added and the incubation continued for an additional 60 min. The molar ratio of double stranded RNA and mRNA is about 200:1. The NOVX mRNA is radiolabeled (using known techniques) and its stability is monitored by gel electrophoresis.

In a parallel experiment made with the same conditions, the double stranded RNA is internally radiolabeled with a <sup>32</sup>P-ATP. Reactions are stopped by the addition of 2 X proteinase K buffer and deproteinized as described previously (Tuschl *et al.*, Genes Dev., 13:3191-3197 (1999)). Products are analyzed by electrophoresis in 15% or 18% polyacrylamide sequencing gels using appropriate RNA standards. By monitoring the gels for radioactivity, the natural production of 10 to 25 nt RNAs from the double stranded RNA can be determined.

The band of double stranded RNA, about 21-23 bps, is eluded. The efficacy of these 21-23 mers for suppressing NOVX transcription is assayed in vitro using the same rabbit reticulocyte assay described above using 50 nanomolar of double stranded 21-23 mer for each assay. The sequence of these 21-23 mers is then determined using standard  
5 nucleic acid sequencing techniques.

### RNA Preparation

21 nt RNAs, based on the sequence determined above, are chemically synthesized using Expedite RNA phosphoramidites and thymidine phosphoramidite (Prologo, Germany). Synthetic oligonucleotides are deprotected and gel-purified (Elbashir,  
10 Lendeckel, & Tuschl, Genes & Dev. 15, 188-200 (2001)), followed by Sep-Pak C18 cartridge (Waters, Milford, Mass., USA) purification (Tuschl, et al., Biochemistry, 32:11658-11668 (1993)).

These RNAs (20  $\mu$ M) single strands are incubated in annealing buffer (100 mM potassium acetate, 30 mM HEPES-KOH at pH 7.4, 2 mM magnesium acetate) for 1 min at  
15 90° C followed by 1 h at 37° C.

### Cell Culture

A cell culture known in the art to regularly express NOVX is propagated using standard conditions. 24 hours before transfection, at approx. 80% confluency, the cells are trypsinized and diluted 1:5 with fresh medium without antibiotics (1-3 X 10<sup>5</sup> cells/ml) and  
20 transferred to 24-well plates (500 ml/well). Transfection is performed using a commercially available lipofection kit and NOVX expression is monitored using standard techniques with positive and negative control. A positive control is cells that naturally express NOVX while a negative control is cells that do not express NOVX. Base-paired 21 and 22 nt siRNAs with overhanging 3' ends mediate efficient sequence-specific mRNA  
25 degradation in lysates and in cell culture. Different concentrations of siRNAs are used. An efficient concentration for suppression in vitro in mammalian culture is between 25 nM to 100 nM final concentration. This indicates that siRNAs are effective at concentrations that are several orders of magnitude below the concentrations applied in conventional antisense or ribozyme gene targeting experiments.

30 The above method provides a way both for the deduction of NOVX siRNA sequence and the use of such siRNA for in vitro suppression. In vivo suppression may be

performed using the same siRNA using well known in vivo transfection or gene therapy transfection techniques.

### Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO:2n, wherein n is an integer between 1 and 52, or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using

chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used). Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-carboxymethylaminomethyl-2-thiouridine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 5-methoxyuracil, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, 2-thiouracil, 4-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a NOVX protein to thereby inhibit expression of the protein (e.g., by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens

expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other. *See, e.g.*, Gaultier, *et al.*, 1987. *Nucl. Acids Res.* **15**: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (*See, e.g.*, Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or a chimeric RNA-DNA analogue (*See, e.g.*, Inoue, *et al.*, 1987. *FEBS Lett.* **215**: 327-330).

#### Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of a NOVX cDNA disclosed herein (*i.e.*, SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 52). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. *See, e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease

activity from a pool of RNA molecules. *See, e.g., Bartel et al., (1993) Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g., the*  
5 NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. *See, e.g., Helene, 1991. Anticancer Drug Des. 6: 569-84; Helene, et al. 1992. Ann. N.Y. Acad. Sci. 660: 27-36; Maher, 1992. Bioassays 14: 807-15.*

In various embodiments, the NOVX nucleic acids can be modified at the base  
10 moiety, sugar moiety or phosphate backbone to improve, *e.g., the stability, hybridization, or solubility of the molecule.* For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g., Hyrup, et al., 1996. Bioorg Med Chem 4: 5-23.* As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g., DNA mimics*) in which the deoxyribose  
15 phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleotide bases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomer can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al., 1996. supra; Perry-O'Keefe, et al., 1996. Proc. Natl.*  
20 *Acad. Sci. USA 93: 14670-14675.*

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g., inducing transcription or translation arrest or inhibiting replication.* PNAs of NOVX can also be used, for example, in the analysis of  
25 single base pair mutations in a gene (*e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S<sub>1</sub> nucleases (See, Hyrup, et al., 1996. supra); or as probes or primers for DNA sequence and hybridization (See, Hyrup, et al., 1996, supra; Perry-O'Keefe, et al., 1996. supra).*

In another embodiment, PNAs of NOVX can be modified, *e.g., to enhance their*  
30 *stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art.* For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g., RNase H and DNA polymerases*) to interact with the DNA

portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleotide bases, and orientation (*see*, Hyrup, *et al.*, 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described  
5 in Hyrup, *et al.*, 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988.  
10 PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*, Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such  
15 as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents  
20 (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

### NOVX Polypeptides

25 A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in any one of SEQ ID NO:2n, wherein n is an integer between 1 and 52. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in any one of SEQ ID NO:2n, wherein n is an integer between 1 and 52,  
30 while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, a NOVX variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other

amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX proteins, still more preferably less than about 10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20%

chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 52) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of a NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of a NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 52. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 52, and retains the functional activity of the protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 52, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 52, and retains the functional activity of the NOVX proteins of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 52.

## **25 Determining Homology Between Two or More Sequences**

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the

molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See*, Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

## 25 Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, a NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a NOVX protein of SEQ ID NO:2n, wherein n is an integer between 1 and 52, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within a NOVX

fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX protein. In one embodiment, a NOVX fusion protein comprises at least one biologically-active portion of a NOVX protein. In another embodiment, a NOVX fusion protein comprises at least two biologically-active portions of a NOVX protein. In yet  
5 another embodiment, a NOVX fusion protein comprises at least three biologically-active portions of a NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

10 In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is a NOVX protein containing a  
15 heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is a NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of  
20 the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a NOVX ligand and a NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of a  
25 NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening  
30 assays to identify molecules that inhibit the interaction of NOVX with a NOVX ligand. A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction

enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.,* Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.,* a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

#### NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.,* mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.,* discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX proteins that function as either NOVX agonists (*i.e.,* mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.,* truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as

individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. See, e.g., Narang, 1983. *Tetrahedron* 39: 3; Itakura, et al., 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, et al., 1984. *Science* 198: 1056; Ike, et al., 1983. *Nucl. Acids Res.* 11: 477.

### Polypeptide Libraries

In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of a NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with  $S_1$  nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the

screening assays to identify NOVX variants. *See, e.g.,* Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

### Anti-NOVX Antibodies

5 Included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, *i.e.*, molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single  
10 chain, F<sub>ab</sub>, F<sub>ab</sub>' and F<sub>(ab)2</sub> fragments, and an F<sub>ab</sub> expression library. In general, antibody molecules obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG<sub>1</sub>, IgG<sub>2</sub>, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to  
15 antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated protein of the invention intended to serve as an antigen, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and  
20 monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 52, and encompasses an epitope thereof such that an  
25 antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly  
30 these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human NOVX protein sequence will

indicate which regions of a NOVX polypeptide are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for  
5 example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each incorporated herein by reference in their entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also  
10 provided herein.

The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three dimensional structural characteristics, as well as specific charge  
15 characteristics. A NOVX polypeptide or a fragment thereof comprises at least one antigenic epitope. An anti-NOVX antibody of the present invention is said to specifically bind to antigen NOVX when the equilibrium binding constant ( $K_D$ ) is  $\leq 1 \mu\text{M}$ , preferably  $\leq 100 \text{ nM}$ , more preferably  $\leq 10 \text{ nM}$ , and most preferably  $\leq 100 \text{ pM}$  to about  $1 \text{ pM}$ , as measured by assays such as radioligand binding assays or similar assays known to those skilled in the  
20 art.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of  
25 polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, *Antibodies: A Laboratory Manual*, Harlow E, and Lane D, 1988; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

### 30 Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by one or more injections with

the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, *etc.*), adjuvants usable in humans such as Bacille Calmette-Guerin and *Corynebacterium parvum*, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (*The Scientist*, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

## 25 Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, *Nature*, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies  
5 that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human  
10 mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell  
15 lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT  
20 medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell  
25 Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. *Immunol.*, 133:3001 (1984); Brodeur et al., *Monoclonal Antibody Production Techniques and Applications*, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

30 The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such

techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). It is an objective, especially important in therapeutic applications of monoclonal antibodies, to identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods (Goding, 1986). Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (*e.g.*, by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

### Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')<sub>2</sub> or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeven et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

### Human Antibodies

Fully human antibodies essentially relate to antibody molecules in which the entire sequence of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be

produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

5           In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, *e.g.*, mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon  
10 challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368,  
15 812-13 (1994)); Fishwild et al. (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the  
20 animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial  
25 chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse<sup>TM</sup> as disclosed in PCT publications WO 96/33735 and WO 96/34096. This  
30 animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be

recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

#### **F<sub>ab</sub> Fragments and Single Chain Antibodies**

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F<sub>ab</sub> expression libraries (see *e.g.*, Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F<sub>ab</sub> fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an F<sub>(ab)<sub>2</sub></sub> fragment produced by pepsin digestion of an antibody molecule; (ii) an F<sub>ab</sub> fragment generated by reducing the disulfide bridges of an F<sub>(ab)<sub>2</sub></sub> fragment; (iii) an F<sub>ab</sub> fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F<sub>v</sub> fragments.

### Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding  
5 target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different  
10 specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO  
15 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least  
20 part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific  
25 antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more  
30 small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or

threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (*e.g.*  $F(ab')_2$  bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, 5 bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate  $F(ab')_2$  fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular 10 disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

15 Additionally, Fab' fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody  $F(ab')_2$  molecule. Each Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was 20 able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies 25 have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. 30 The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain ( $V_H$ ) connected to a light-chain variable domain ( $V_L$ ) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the  $V_H$  and  $V_L$  domains

of one fragment are forced to pair with the complementary V<sub>L</sub> and V<sub>H</sub> domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

5       Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a  
10       triggering molecule on a leukocyte such as a T-cell receptor molecule (*e.g.* CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells  
15       which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

#### **Heteroconjugate Antibodies**

Heteroconjugate antibodies are also within the scope of the present invention.  
20       Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies, for example, target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example,  
25       immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

#### **Effector Function Engineering**

30       It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, *e.g.*, the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing

interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

#### 10 **Immunoconjugates**

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (*e.g.*, an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*i.e.*, a radioconjugate).

15 Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include  $^{212}\text{Bi}$ ,  $^{131}\text{I}$ ,  $^{131}\text{In}$ ,  $^{90}\text{Y}$ , and  $^{186}\text{Re}$ .

25 Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled

1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

5 In another embodiment, the antibody can be conjugated to a "receptor" (such as streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (*e.g.*, avidin) that is in turn conjugated to a cytotoxic agent.

### Immunoliposomes

10 The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein *et al.*, *Proc. Natl. Acad. Sci. USA*, 82: 3688 (1985); Hwang *et al.*, *Proc. Natl. Acad. Sci. USA*, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

15 Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in  
20 Martin *et al.*, *J. Biol. Chem.*, 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon *et al.*, *J. National Cancer Inst.*, 81(19): 1484 (1989).

### Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention

In one embodiment, methods for the screening of antibodies that possess the desired  
25 specificity include, but are not limited to, enzyme linked immunosorbent assay (ELISA) and other immunologically mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an NOVX protein is facilitated by generation of hybridomas that bind to the fragment of an NOVX protein possessing such a domain. Thus, antibodies that are specific for a desired domain  
30 within an NOVX protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Antibodies directed against a NOVX protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of a NOVX protein (e.g., for use in measuring levels of the NOVX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies specific to a NOVX protein, or derivative, fragment, analog or homolog thereof, that contain the antibody derived antigen binding domain, are utilized as pharmacologically active compounds (referred to hereinafter as "Therapeutics").

An antibody specific for a NOVX protein of the invention (e.g., a monoclonal antibody or a polyclonal antibody) can be used to isolate a NOVX polypeptide by standard techniques, such as immunoaffinity, chromatography or immunoprecipitation. An antibody to a NOVX polypeptide can facilitate the purification of a natural NOVX antigen from cells, or of a recombinantly produced NOVX antigen expressed in host cells. Moreover, such an anti-NOVX antibody can be used to detect the antigenic NOVX protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic NOVX protein. Antibodies directed against a NOVX protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

### 30 Antibody Therapeutics

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may be used as therapeutic agents. Such agents will generally be employed

to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand which may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume other subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

### Pharmaceutical Compositions of Antibodies

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed. (Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995;

Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York.

If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, e.g., Marasco et al., Proc. Natl. Acad. Sci. USA, 90: 7889-7893 (1993). The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended.

The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

The formulations to be used for in vivo administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g., films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and  $\gamma$  ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT<sup>TM</sup> (injectable microspheres composed of

lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

## 5 ELISA Assay

An agent for detecting an analyte protein is an antibody capable of binding to an analyte protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g.,  $F_{ab}$  or  $F_{(ab)2}$ ) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. Included within the usage of the term "biological sample", therefore, is blood and a fraction or component of blood including blood serum, blood plasma, or lymph. That is, the detection method of the invention can be used to detect an analyte mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of an analyte mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of an analyte protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of an analyte genomic DNA include Southern hybridizations. Procedures for conducting immunoassays are described, for example in "ELISA: Theory and Practice: Methods in Molecular Biology", Vol. 42, J. R. Crowther (Ed.) Human Press, Totowa, NJ, 1995; "Immunoassay", E. Diamandis and T. Christopoulos, Academic Press, Inc., San Diego, CA, 1996; and "Practice and Theory of Enzyme Immunoassays", P. Tijssen, Elsevier Science Publishers, Amsterdam, 1985. Furthermore, *in vivo* techniques for detection of an analyte protein include introducing into a subject a labeled anti-analyte protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

### NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory

sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., NOVX proteins, mutant forms of NOVX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION

TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. See, e.g., Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (see, e.g., Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1:

268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.*, Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

5           Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or  
10   electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

          For stable transfection of mammalian cells, it is known that, depending upon the  
15   expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and  
20   methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

25           A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable  
30   medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

### Transgenic NOVX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, *etc.* A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences, *i.e.*, any one of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and

microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other  
5 transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying  
10 other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of any one of SEQ ID NO:2*n*-1, wherein *n* is an  
15 integer between 1 and 52), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 52, can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous  
20 recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby  
25 alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient  
30 length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.*, Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is then introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced NOVX gene has

homologously-recombined with the endogenous NOVX gene are selected. *See, e.g., Li, et al., 1992. Cell 69: 915.*

The selected cells are then injected into a blastocyst of an animal (*e.g., a mouse*) to form aggregation chimeras. *See, e.g., Bradley, 1987. In: TERATOCARCINOMAS AND*  
5 *EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp.*  
113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the  
homologously-recombined DNA in their germ cells can be used to breed animals in which  
all cells of the animal contain the homologously-recombined DNA by germline  
10 transmission of the transgene. Methods for constructing homologous recombination  
vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr.*  
*Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO  
91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that  
15 contain selected systems that allow for regulated expression of the transgene. One example  
of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description  
of the cre/loxP recombinase system, *See, e.g., Lakso, et al., 1992. Proc. Natl. Acad. Sci.*  
*USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase  
system of *Saccharomyces cerevisiae*. *See, O'Gorman, et al., 1991. Science* 251:1351-1355.  
20 If a cre/loxP recombinase system is used to regulate expression of the transgene, animals  
containing transgenes encoding both the Cre recombinase and a selected protein are  
required. Such animals can be provided through the construction of "double" transgenic  
animals, *e.g., by mating two transgenic animals, one containing a transgene encoding a*  
selected protein and the other containing a transgene encoding a recombinase.

25 Clones of the non-human transgenic animals described herein can also be produced  
according to the methods described in Wilmut, *et al., 1997. Nature* 385: 810-813. In brief,  
a cell (*e.g., a somatic cell*) from the transgenic animal can be isolated and induced to exit  
the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be fused, *e.g., through the*  
use of electrical pulses, to an enucleated oocyte from an animal of the same species from  
30 which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it  
develops to morula or blastocyte and then transferred to pseudopregnant female foster  
animal. The offspring borne of this female foster animal will be a clone of the animal from  
which the cell (*e.g., the somatic cell*) is isolated.

### Pharmaceutical Compositions

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions  
5 suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are  
10 described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for  
15 pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with  
20 its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols,  
25 glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium  
30 hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous

preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable

5 under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a

10 coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol,

15 sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a NOVX protein or anti-NOVX antibody) in the required amount in an appropriate

20 solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and

25 freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and

30 used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The

tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such  
5 as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

10 Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of  
15 nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

20 In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation  
25 of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described  
30 in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to

produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding  
5 such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91:  
10 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery  
15 system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

#### Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express NOVX  
20 protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.*, in a biological sample) or a genetic lesion in a NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or  
25 excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious  
30 disease(possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention can be used in methods

to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

## 5 Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein  
10 activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a NOVX protein or polypeptide or biologically-active portion thereof. The test compounds  
15 of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is  
20 limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small  
25 molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the  
30 art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed.*

*Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (e.g., Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

10        In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a NOVX protein determined. The cell, for example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be  
15        accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^3\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or  
20        by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a  
25        known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known  
30        compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the NOVX protein or

biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule. As used herein, a "target molecule" is a molecule with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a NOVX protein or polypeptide of the invention. In one embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular  $\text{Ca}^{2+}$ , diacylglycerol,  $\text{IP}_3$ , *etc.*), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (comprising a NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to

interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

5 In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (e.g. stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by  
10 determining the ability of the NOVX protein to bind to a NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate a NOVX target molecule. For example, the catalytic/enzymatic activity of the target  
15 molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein  
20 determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of a NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the  
25 membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton<sup>®</sup> X-100, Triton<sup>®</sup> X-114, Thesit<sup>®</sup>,  
30 Isotridecypoly(ethylene glycol ether)<sub>n</sub>, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate

separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX

mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

### Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the  
5 corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. Some of these applications  
10 are described in the subsections, below.

### Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences  
15 of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52, or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers  
20 (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the NOVX  
25 sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme,  
30 but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or

a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. *See, e.g., D'Eustachio, et al., 1983. Science 220: 919-924.* Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with  
5 translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific  
10 chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes  
15 can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection.  
20 Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see, Verma, et al., HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES* (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for  
25 marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical  
30 position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g., in McKusick, MENDELIAN INHERITANCE IN MAN*, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through

linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the NOVX gene, can be determined. If a  
5 mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence.  
10 Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

#### Tissue Typing

The NOVX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested  
15 with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative  
20 technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner,  
25 can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater  
30 degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If coding sequences, such as those of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 52, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

## 10 Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in a NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or

prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents  
5 (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

These and other agents are described in further detail in the following sections.

### Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting  
10 the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX  
15 nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 52, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

20 An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking)  
25 a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include  
30 tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern

hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

## 25 Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression

or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in a NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (i) a deletion of one or more nucleotides from a NOVX gene; (ii) an addition of one or more nucleotides to a NOVX gene; (iii) a substitution of one or more nucleotides of a NOVX gene, (iv) a chromosomal rearrangement of a NOVX gene; (v) an alteration in the level of a messenger RNA transcript of a NOVX gene, (vi) aberrant modification of a NOVX gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a NOVX gene, (viii) a non-wild-type level of a NOVX protein, (ix) allelic loss of a NOVX gene, and (x) inappropriate post-translational modification of a NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a NOVX gene. A preferred biological sample is a peripheral blood

leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see*, Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see*, Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see*, Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Q $\beta$  Replicase (*see*, Lizardi, *et al.*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes

(see, e.g., U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high-density arrays containing  
5 hundreds or thousands of oligonucleotide probes. See, e.g., Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to  
10 identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one  
15 complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim  
20 and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (see, e.g., Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36:  
25 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. See, e.g., Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by  
30 hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids

treated with S<sub>1</sub> nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated  
5 by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.,* Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more  
10 proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662.  
15 According to an exemplary embodiment, a probe based on a NOVX sequence, *e.g.,* a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

20 In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX  
25 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may  
30 be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g., Myers, et al., 1985. Nature 313: 495.* When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g., Rosenbaum and Reissner, 1987. Biophys. Chem. 265: 12753.*

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g., Saiki, et al., 1986. Nature 324: 163; Saiki, et al., 1989. Proc. Natl. Acad. Sci. USA 86: 6230.* Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g., Gibbs, et al., 1989. Nucl. Acids Res. 17: 2437-2448*) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g., Prossner, 1993. Tibtech. 11: 238*). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g., Gasparini, et al., 1992. Mol. Cell Probes 6: 1.* It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g., Barany, 1991. Proc. Natl. Acad. Sci. USA 88: 189.* In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.,* in clinical settings to diagnose

patients exhibiting symptoms or family history of a disease or illness involving a NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein.

- 5 However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

### Pharmacogenomics

- Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (e.g., NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders. The disorders include but are not limited to, e.g., those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

- In conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

- Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See e.g., Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example,

glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

5       As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome pregnancy zone protein precursor enzymes CYP2D6 and CYP2C19) has  
10       provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the  
15       absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid  
20       metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

      Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition,  
25       pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a NOVX modulator, such as a modulator identified by one of  
30       the exemplary screening assays described herein.

### Monitoring of Effects During Clinical Trials

      Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or

differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a NOVX protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample

with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

### Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

These methods of treatment will be discussed more fully, below.

### Diseases and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (i) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (ii) antibodies to an aforementioned peptide; (iii) nucleic acids encoding an aforementioned peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with

Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

- 5 Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by
- 10 sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, *etc.*) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

### Prophylactic Methods

- In one aspect, the invention provides a method for preventing, in a subject, a disease
- 15 or condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur
- 20 prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, a NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are
- 25 further discussed in the following subsections.

### Therapeutic Methods

- Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of
- 30 NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a NOVX protein, a peptide, a NOVX

peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples  
5 of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a NOVX protein or nucleic acid  
10 molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering a NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

15 Stimulation of NOVX activity is desirable *in situations* in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (e.g., cancer or immune associated disorders). Another example of such a situation is where the subject has a  
20 gestational disease (e.g., preclampsia).

#### **Determination of the Biological Effect of the Therapeutic**

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

25 In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly,  
30 for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

## Prophylactic and Therapeutic Uses of the Compositions of the Invention

The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from diseases, disorders, conditions and the like, including but not limited to those listed herein.

Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies, which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

## EXAMPLES

### Example A: Polynucleotide and Polypeptide Sequences, and Homology Data

#### Example 1.

The NOV1 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 1A.

Table 1A. NOV1 Sequence Analysis			
	SEQ ID NO: 1	2566 bp	
NOV1a, CG108030-01 DNA Sequence	GGNCACGAGCGGCCCTCCACTCCCTGACTGTCGTGTTTGTCTCGCTCTGTGCTGAGGGCTGATG		
	CTGAGGACCTCCTTGACTCCTTCCTTAGCAACATTCTACAGGACTGCAGGCACCACCTGTGTGA		
	ACCGGACATGAACTGGTGTGGCCTAGTGCCAAGCTGTTGCAGGCAGCTGCAGGTGCATCTGCC		
	CGGGCCTGTGACTCTGTCAACGCAAGTACTGCCTTTACTGCTGGAACAGTTCCACAAGCACAG		
	TCAGAGCAGCCAGCGCGGACAATCCTTGAAATGCTCCTGGGTTTCTTGAAGCTGCAGCAGAAA		
	TGGAGCTATGAAGACAAAGATCAAAGGCCTCTGAATGGCTTCAAGGACCAGCTGTGCTCACTGG		
	TATTCATGGCTCTAACAGACCCCGACCCAGCTTCAGCTTGTGGCATCCGTACACTGACAGT		

	SEQ ID NO: 2	389 aa	MW at 42642.8kD
NOV1a, CG108030-01 Protein Sequence	MLLGFLKLQQKWSYEDKDQRPINGFKDQLCSLVFMALTDPSLQQLVGIRITLTVLGAQPDLLSY EDLELAVGHLYRLSPFKEDSQSCRVAALASGTLAALYPVAFSSHLPKLAEEELRVGESNLITNG DEPTQCSRHLCCQLALSAVSTHPSIVKETPLLLQLHLWQVNRGNMVAQSSDVIACVQSLRQMAE KCQQDPESCWFYHQTAIPCLLALAVQASMPKEKPSVLKRVLEDEVLAAMVSVIGTATTLLSPE LAAQSVTHIVPLFLDGNVSFLPENSFPSPRFQPFQDGSSGQRRILALLMAFVCSLPRNVEIPQLN QLMRELLLELSCCHSCFPSSSTAACCFAGLLNKHYPACQQLDEFLQLAVDKVEAGLAGLPVVVRPS LFFSG		

	SEQ ID NO: 3	3319 bp
NOV1b, CG108030-02 DNA Sequence	TCGCGTTATGGCCGCTGCCGCGGCTGTGAGGCGGGCGCCTATGGGTGCCCTATGGGGCCTC GTGCACGACTTCGTCGTGGGTGAGCAAGAGGGCCCCGCTGACCAGGTGGCTGCAGATGTGAAAT CTGCAACATATACAGTGTTCACAGTTGTGGAAGCCCTTGGGTCTCTCTAGAGAAATCCAGAAC CCGAATCTCGGGCAGGAGAAATCCAGCTTTTGTACAGGTGCTACTCCACTGTCAACCTTGCTC CTGGAGAAGGAAGTGGTACCTGTATCTGTTATGAGAACCGGCTGAAGGACCATCATCTTG TGATCCCATCTGTCTGTCAGGGTTTGAAGGCACTTAGCCCTGTGTGTGGCCCTGCCCCAGGGCT GGCTGTTTTCTGTGCTTAAAGCCATCTTCCAGGAAGTGCATGTACAGTCCCTGCCACAGGTGGAC CGACACACAGTCTACAATATCATCACTTAATTTATGCGAACCCGGGAAGAAGAGCTAAAGAGCC TAGGAGCTGACTTCACCTTTGGCTTATCCAGGTGATGGATGGGGAAGAGATCCCCGTAATCT TCTGTGGCCTTCCGCATCGTCCATGACCTCATCTCCAGGACTATAGCCTGGGACCTTTGTG GAGGAGTTGTTGAAAGTGACATCTGTTATTTCCCTATCGATTTACCCCTCCACTTAATGTC CCCATGGTATCCAGAGGAAGACCTCATCTGAGCTTCCGCTGTGTGGCTTCTACACCACTG ATTGCTGAGTTTCTGCTGCCCCCTGTTGATTGAGAAAGTGGATTCTGAGGTTCTGAGTGCCAAG TTGGATTCTCTACAGACTCTGAATGCTTGTGTGCTGTGTATGACACAGGAAGAACTGAAGGACT TCTCTCCCAGCCTTTGGGCTTCTATCCGAGAGAGGTTTCCAGACGGCAAGTGAGCGGGTGA GGCAGAGGGCTGGCGGCCCTCCACTCCCTGACTGCGGTGTTTGTCTCGCTCTGTGCTGAGGGC	

	<p>GATGCTGAGGACCTCCTTGACTCCTTCCTTAGCAACATTCTACAGGACTGCAGGCACCACCTGT  GTGAACCGGACATGAAACTGGTGTGGCCTTAGTGCAAGCTGTTGCAGGCAGCTGCAGGTGCATCT  GCCCCGGCCTGTGACTCTGTCAACAGCAATGTACTGCCTTTACTGCTGGAACAGTTCCACAAGC  ACAGTCAGAGCAGCCAGCGGCGGACAAATCCTTGAAATGCTCCTGGGTTTCTTGAACTGCAGC  AGAAATGGAGCTATGAAGACAAAGATCAAAGGCCTCTGAATGGCTTCAAGGACCAGCTGTGCTC  ACTGGTATTCATGGCTCTAAAGACCCCGACCCAGCTTCAGCTTGTGGCATCCGTACACTG  ACAGTCTTGGGTGCCCAGCCAGATCTCCTATCTTATGAGGACTTGGAGCTGGCAGTGGGTCAACC  TGTACAGACTGAGCTTCTGAAGGAGGATTCCAGAGTTGCAGGGTGGCAGCACTGGAAGCATC  AGGAACCTTGGCTGCTCTACCTGTGGCCTTCAGCAGCCACCTCGTACCCAAGCTCGCTGAG  GAGCTGCGTGTAGGGGAGTCAAATTTGACTAACGGAGATGAGCCACCCAATGCTCCCGGCATC  TGTGCTGTCTGCAAGCCTTGTCACTGTATCAACACATCCAGCATCGTCAAGGAGACTGCCC  TCTGCTGTCTGAGCATCTCTGGCAAGTGAACAGAGGGAATATGGTTGCAATCCAGTGACGTT  ATTGCTGTCTGTCAAGCCTCAGACAGATGGCAGAAAAATGTGAGCAGGACCCTGAGAGTTGCT  GGTATTTCCACCAGACAGCTATACCTTGCCTTGCCTTGGCTTGGCTGTGAGCAGCTCTATGCCCAG  GAAGGAGCCCTCAGTTCTGAGAAAAGTACTATTGGAGGATGAGGTGTTGGCTGCCATGGTGTCT  GTCAATTGGCACTGTACAAACCCACTGAGCCCTGAGTTAGCTGCCAGAGTGTGACACACATTG  TGCCCTCTTCTTGGATGGCAACGTGTCTTCTGCTGAAAACAGCTTCCGAGCAGATTCCA  GCCATTCCAGGATGGCTCCTCAGGCGAGAGCGGCTGATTGCACTGCTTATGGCCTTGTCTGC  TCCCTGCCTCGAAATGGCAGCAGCTGGATGAATTCTACAGCTAGCTGTGGACAAAGTGGAGGC  TGGCCTGGACTCTGGGCCCTGTCTGAGTCAAGCCCTCACTCTTCTCTGGGTAAACAAAGGCC  CTAGTGTCTCAGATACCATCTCTCAGCTCCTGCCTTACAGCCCGCTCATGGCCCTCTGAGT  ACCCAGAATTAGGTCCAGCAGCAGCTGATGGCTTCTCTCTGCTCATGTCTGACTGCACTGATGT  GCTGACTCGTCTGGCCATGCCGAAGTGGCGATCATGTTCCGCCAGCGGTTCTTACAGATAAT  GTGCTGCTTGGTCCAGACTTCCATGCTGCTCCCCAAGATGTGAAGCCAACTACTTGAAG  GTCTTTCTCATGTACTTAACAGGCTGCCCAAGCCTGTACTCTTGCCAGAGCTGCCCAAGCTTCT  TTCCTTGTCTGAGGAGCCCTGTCTGCTGCTGAGTGTGTGGTGCAGCTCTCCACCCTCAGCTGC  CTTCAAGCCTCTTCTACTGAAGCAACCCCAAGTCACTGAGTCTTCACTGGAACCCCTCGTCAACA  AGTTTCTGAACCTCAGCTCTAGCCCTTCCATGGCTGTCCGGATCGCCGCACTGCAGTGCATGCA  TGCTCTCACTCGCCTGCCACCCCTGTGCTGCTGCCGTACAAACCAAGGTGATTGGGCCCTTA  GCCAAACCCCTGGATGACAGAAGAGACTGGTGCAGGAAGCAGTGTGACCCAGAGGGGAGT  GGTTTCTGTTGGGAGCCCTGGCAGCTGAGCCCTCAGTCTTGGCTAGACTGTTTGCACAATCT  AACCTGGGATTACTAATGTTGAGCCATCTTCCCAAGCAGGGAACCACTGGTCTCTGACTG  CCTTCCCACAGACACAGCAAAATGCTAGGCCTCTGTTGATGGCTGTACAAAGACATAAGA  GTCCATATTTCTAGTGGATTGTAAAATAAGTGTGTGTGAGACACTTGCCTTTGAAGAAAGATC  TAGGGTCTTGGGTCTCTTGCATTATATGTGAGAAAGGGGCGATATGCTGCTGAGGGGTGAGT  GCATATGAGTGTGGCCCTGAGGACAGGGCTGGCAGATGTTGTCTACCTGCTGAG</p>
	<p>ORF Start: ATG at 8</p>
	<p>ORF Stop: TAG at 2219</p>

NOV1b, CG108030-02 Protein Sequence	<p>SEQ ID NO: 4</p> <p>737 aa</p> <p>MW at 81317.6kD</p> <p>MAAAAVEAAAPMGALWGLVHDFVVGQOQEGPADQVAADVKSNGYTVLQVVEALGSSLENPEPT  RARGIQLLSQVLLHCHTLLLEKEVHLLIFYENRLKDHHLVIPSVLQGLKALSCLVALPPGLAV  SVLKAIQFQEVHVQSLPQVDRHTVYNIITNFMRTREEELKSLGADFTFGFIQVMDGEKDPNNLLV  AFRIVHDLISRDSYSLGPFVEELFEVTSYFPIIDFTPPNDPHGQIREDLILSLRAVLASTPFA  EFLPLLLIEKVDSEVLSAKLDSLQTLNACCAVYGQKELKDFLPSLWASIRREVFTASERVEAE  GLAALHSLTACLSRSVLRADAEDLLDSFLSNILQDCRHLCEPDMKLVWPSASCCRLQVHLPG  PVTLSAMYCLYCWNSSSTVRAASGGTILEMLLGFLKLQKWSYEDKQRPNGFKDQLCSLV  FMALTDPSLQLQVGIKTLTVLGAQPDLLSYEDLELAVGHLVRLSFLKEDSQSCRVALEASGT  LAALYPVAFSSHLVPLKAEELRVGESNLNGDEPTQCSRHLCLQALSAVSTHPSIVKETLPL  LQHLWQVNRGNMVAQSSDVIAVCQSLRQMAEKQODPESCWYFHQTAIPCLLALAVQASMPEKE  PSVLRKVLLEDEVLAAMVSVIGTATTHLSPELAAQSVTHIVPLDGNVSFLPENSFPSPRFQPF  QDSSGQRRLLALLMAFVCSLPRNGSSWMNSYS</p>
---	---

5 Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 1B.

Protein Sequence	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region
------------------	-----------------------------------	--

NOV1b	1..313 416..728	313/313 (100%) 313/313 (100%)
-------	--------------------	----------------------------------

Further analysis of the NOV1a protein yielded the following properties shown in Table 1C.

Table 1C. Protein Sequence Properties NOV1a	
SignalP analysis:	No Known Signal Sequence Indicated
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 11; pos.chg 2; neg.chg 0  H-region: length 3; peak value -19.72  PSG score: -24.12</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): -3.97  possible cleavage site: between 40 and 41</p> <p>&gt;&gt;&gt; Seems to have no N-terminal signal peptide</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 1  Tentative number of TMS(s) for the threshold 0.5:  0  number of TMS(s) .. fixed  PERIPHERAL Likelihood = 0.58 (at 232)  ALOM score: 0.58 (number of TMSs: 0)</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 0 Hyd Moment(75): 4.03  Hyd Moment(95): 7.41 G content: 1  D/E content: 2 S/T content: 1  Score: -7.36</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  cleavage site motif not found</p> <p>NUCDISC: discrimination of nuclear localization signals  pat4: none  pat7: none  bipartite: none  content of basic residues: 7.2%  NLS Score: -0.47</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals: none</p> <p>SKL: peroxisomal targeting signal in the C-terminus:  none</p> <p>PTS2: 2nd peroxisomal targeting signal: none</p>

	<p>VAC: possible vacuolar targeting motif: none</p> <p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif: type 1: none type 2: none</p> <p>NMYR: N-myristoylation pattern : none</p> <p>Prenylation motif: none</p> <p>memYQRL: transport motif from cell surface to Golgi: none</p> <p>Tyrosines in the tail: none</p> <p>Dileucine motif in the tail: none</p> <p>checking 63 PROSITE DNA binding motifs: none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p> <p>checking 33 PROSITE prokaryotic DNA binding motifs: none</p> <p>NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination Indication: cytoplasmic Reliability: 76.7</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions total: 0 residues</p> <p>-----</p> <p>Final Results (k = 9/23):</p> <p>47.8 %: nuclear 26.1 %: cytoplasmic 17.4 %: mitochondrial 4.3 %: vacuolar 4.3 %: vesicles of secretory system</p> <p>&gt;&gt; indication for CG108030-01 is nuc (k=23)</p>
--	--

A search of the NOV1a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 1D.

Table 1D. Geneseq Results for NOV1a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV1a Residues/	Identities/ Similarities for	Expect Value

		Residues	Region	
AAB61304	Human transcriptional regulator protein #4 - Homo sapiens, 615 aa. [WO200078954-A2, 28-DEC-2000]	1..378 1..378	376/378 (99%) 376/378 (99%)	0.0
AAU28025	Novel human secretory protein, Seq ID No 194 - Homo sapiens, 666 aa. [WO200166689-A2, 13-SEP-2001]	1..378 52..429	376/378 (99%) 376/378 (99%)	0.0
AAB93270	Human protein sequence SEQ ID NO:12306 - Homo sapiens, 774 aa. [EP1074617-A2, 07-FEB-2001]	1..378 160..537	375/378 (99%) 375/378 (99%)	0.0
AAM41729	Human polypeptide SEQ ID NO 6660 - Homo sapiens, 398 aa. [WO200153312-A1, 26-JUL-2001]	1..314 67..380	314/314 (100%) 314/314 (100%)	e-180
AAM39943	Human polypeptide SEQ ID NO 3088 - Homo sapiens, 383 aa. [WO200153312-A1, 26-JUL-2001]	1..314 52..365	314/314 (100%) 314/314 (100%)	e-180

In a BLAST search of public sequence databases, the NOV1a protein was found to have homology to the proteins shown in the BLASTP data in Table 1E.

Table 1E. Public BLASTP Results for NOV1a				
Protein Accession Number	Protein/Organism/Length	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96T76	MMS19 - Homo sapiens (Human), 1030 aa.	1..378 416..793	376/378 (99%) 376/378 (99%)	0.0
Q9BUE2	Hypothetical protein - Homo sapiens (Human), 692 aa (fragment).	1..378 78..455	376/378 (99%) 376/378 (99%)	0.0
Q9BYS9	MMS19 protein - Homo sapiens (Human), 1030 aa.	1..378 416..793	376/378 (99%) 376/378 (99%)	0.0
Q96DF1	MMS19 (MET18 <i>S. cerevisiae</i> )-like - Homo sapiens (Human), 666 aa.	1..378 52..429	376/378 (99%) 376/378 (99%)	0.0
Q96RK5	Transcriptional coactivator MMS19 - Homo sapiens (Human), 1030 aa.	1..378 416..793	375/378 (99%) 375/378 (99%)	0.0

5

### Example 2.

The NOV2 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 2A.

Table 2A. NOV2 Sequence Analysis			
	SEQ ID NO: 5	3058 bp	
NOV2a, CG115907-01 DNA Sequence	GCCCCACAGTGAGAGGAAGGAAGGCAACAGTCGCCAGCAGCCGATGTGAAGACCGGACTCCGTG CGCCCCCTCGCCGCTCTGCTGGCCACATCGATGTTGTGTCCGCCGCTGCTCGCCCCGATCAC GATGAAGCCCCCAAGGCCTGTCCGTACCTGCAGCAAAGTTCTCGTCTGCTTTCACTGCTGGCC ATCCACCAGACTACTACTGCCGAAAAGATGGCATCGACATCTACAGCCTCACCGTGGACTCCA GGGTCTCATCCCCGATTTGCCCAACCGGTGTCACAGCCGAGTGGTCAATAGGGCCAATACTGT GCAGGAGGCCACCTTCCAGATGGAGCTGCCAAGAAAGCCTTCATCAACCACTTCTCCATGATC ATCGATGGCATGACCTACCCAGGGATCATCAAGGAGAAGGCTGAAGCCCAGGCACAGTACAGCG CAGCAGTGGCCAAGGGAAGAGCGCTGGCCTCGTCAAGGCCACCGGGAGAAAATGGAGCAGTT CCAGGTGTCGGTCACTGTGGCTCCCAATGCCAAGATCACCTTTGAGCTGGTCTATGAGGAGCTG CTCAAGCGGCGTTTGGGGGTGTACGAGCTGCTGTGAAAGTGCAGGCCAGCAGCTGGTCAAGC ACCTGCAGATGGACATTCACATCTTCGAGCCCCAGGGCATCAGCTTCTGAGAGACAGAGAGCAC CTTCATGACCAACAGCTGGTAGACGCCCTCACACCTGGCAGAATAAGACCAAGGCTCACATC CGGTTCAGCCAAACACTTTCACAGCAGCAAAGTCCCCAGAGCAGCAAGAAACAGTCTGGACG GCAACCTCATTATCCGCTATGATGTGGACCCGGGCCATCTCCGGGGCTCCATTCAGATCGAGAA CGGCTACTTTGTACTACTTTGCCCCGAGGGCCTAACCAAAATGCCAAGAATGTGGTCTTT GTCATTGACAAGAGCGGCTCCATGAGTGGCAGGAAAATCCAGCAGACCCGGGAAGCCCTAATCA AGATCCTGGATGACCTCAGCCCCAGAGACAGTTCAACCTCATCGTCTTCAGTACAGAAGCAAC TCAGTGGAGGCCATCAGTGTGCCAGCCTCAGCCGAGAACGTGAACAAGGCCAGGAGCTTTGCT GCGGGCATCCAGGCCCTGGGAGGGACCAACATCAATGATGCAATGCTGATGGCTGTGAGTTGC TGGAACAGCAGCAACAGGAGGAGCGGCTGCCGAAGGGAGTGTCTCACTCATCATCTGCTCAC CGATGGCGACCCCACTGTGGGGAGACTAACCCAGGAGCATCCAGAATAACGTGCGGGAAGCT GTAAGTGGCCGGTACAGCCTCTTCTGCTGGGCTTCGGTTTCGAGCTAGCTATGCCTTCCTGG AGAAGCTGGCACTGGACAATGGCGGCTGGCCCGGCGCATCCATGAGGACTCAGACTCTGCCCT GCAGCTCCAGGACTTCTACAGGAAGTGGCCAACCCACTGCTGACAGCAGTGAACCTTCGAGTAC CCAAGCAATGCCGTGGAGGAGGTCACTCAGAACAACTTCGGCTCTCTTCAAGGGCTCAGAGA TGGTGGTGGCTGGGAAGCTCCAGGACCGGGGGCTGATGTGCTCAGGCCAGTCACTGGGAA GCTGCCCTACACAGAACATCACTTCCAAACGAGTCCAGTGTGGCAGAGCAGGAGGCGGAGTTC CAGAGCCCCAAGTATATCTTCCACAACCTTCATGGAGAGGCTCTGGGCATACCTGACTATCCAGC AGCTGCTGGAGCAAACTGTCTCCGCATCCGATGCTGATCAGCAGGCCCTCCGGAACCAAGCGCT GAATTTATCACTTGCCCTACAGCTTTGTACGCCTCTCACATCTATGGTAGTACCAAACCCGAT GACCAAGAGCAGTCTCAAGTTGCTGAGAAGCCCATGGAAGGCGAAAGTAGAAAACAGGAATGTCC ACTCAGGTTCCACTTTCTTCAAATATTATCTCCAGGGAGCAAAAATACAAAACAGAGGCTTC CTTTTCTCCAAGAAGAGGATGGAATAGACAAGCTGGAGCTGCTGGCTCCCGATGAATTTCAGA CTTGGGGTTCTCAGCTCCAGGCAACTTGGACTCCAGGACCTCTGATGTTCTGACCATGCTG CTTACCACCCCTTCCGCCGCTCTGGCCATCTTGCTGCTTCAGCACCACAGCCACCTCAAATCC TGATCCAGCTGTGTCTCGTGTATGAATATGAAAATCGAAGAAACCAATGACACCCCAAACC CCAGCCCCCATAAGGCTCCCTCTGCCATCTGCTGCTGCTGGCAGAGTGTGGAGCGGCTCT GTGTGGACCCCAAGACACCGCCAGGGGCCAGTGAACCTGCTCTCAGACCTGAGCAAGGGTTGA GGTGACTGGCCAGTATGAGAGGGAGAAGGCTGGGTTCTCATGGATCGAAGTGACCTTCAAGAAC CCCCTGGTATGGGTTCAAGCATCCCTGAACACGTGGTGGTGAATCGGAACCCGAAGAGCTCTG CGTACAAGTGAAGGAGAGCGCTATTCTCAGTATGCCCCGCTGAAGATGACCATGGACAAGAC GGGTCTCTGCTGCTCAGTGACCCAGACAAAGTGACCATCGGCCCTGTTGTTCTGGGATGGCCGT GGGGAGGGCTCCGGCTCCTTCTGCGTGACACTGACCGCTTCTCCAGCCAGGTTGGAGGGACCC TTGGCCAGTTTTACAGGAGGTGCTCTGGGGATCTCCAGCAGATCAGATGACGGCAGACGCAC GCTGAGGGTTCAAGGCAATGACCACTCTGCCACCAGAGAGCGCAGGCTGGATTACAGGAGGGG CCCCGGGAGTGGAGATTCTCTGCTGGTCTGTGGAGCTGTAGTTCTGATGGAAGGAGCTGTGCC CACCCTGTACACTTGGCTTCCCCCTGCAACTGCAGGGCCGCTTCTGGGGCTGGACCAACATGG GGAGGAAGAGTCCCACTATTACAATAAAGAAAGGTGGTGTGAGCTGA		
	ORF Start: ATG at 130		ORF Stop: TAG at 2920

	SEQ ID NO: 6	930 aa	MW at 103356.4kD
NOV2a, CG115907-01 Protein Sequence	MKPPRPVRTCSKVLVLLSLLAIHQTTAEKNGIDIYSLTVDSRVSSRFAHTVTVTSRVVNRANTV QEATFQMEPLPKAFITNFSMIIDGMTYPGIIKEKAEQAQYSAVAKGSAGLYKATGRNMEQF QVSVSVAPNAKITFELVYEEELKRRRLGVYELLKVRPQQLVKHLQMDIHIFEPQGISFLETEST FMTNQLVDALTWQNKTKAHIRFKPTLSQQQKSPEQQETVLDGNLIIRYDVDRASGGSIQIEN GYFVHYFAPEGLTTPKNNVVFIDKSGSMSGRKIQQTREALIKILDDLSPRDQFNLI V FST EAT QWRPSLVPAEENVNKARSFAAGTQALGGTNINDAMLMAVQLLDSNQEERLPEGSVSLI ILLT DGDPTVGETNPRSIQNNVREAVSGRYSFLGFGFDVSYAFLEKLALDNGGLARRIHEDSDSAL QLQDFYQEVANPLLTAVTFEYPSNAVEEVTQNNFRLLFKGSEM V VAGKLQDRGPDVLTATVSGK LPTQNIITFQTESSVABQEAEPQSPKYPHNFMERLWAYLTIQQLLEQTVSASDADQALRNQAL NLSLAYSFVTPLTSMVVTKPDDQEQSQA EKPMEGESRNRNVHSGSTFFKYLLQGAKIPKPEAS FSPRRGWNROGAGAAGSRMNFPRGVLSSRQLGLPGPPDVPDHAAYHPFRRLAILPASAPPATSNP DPAVSRVMNMKIEETMTTQTPAPIQAPSAILPLPGQSVERLCVDPHRHQGPVNLLSDPEQGV E VTGQYEREKAGFSWIEVTFKNPLVWVHASPEHVVTNRNRSSAYKWKETLFSVMPGLKMTMDKT GLLLLSDDPKVTIGLLFWDGRGEGRLRLRLDRDTRFSSHVGGTLGQFYQEV LWGSPAASDDGRRT LRVQGNDSATRERRLDYQEGPPGVEISCWSVEL		

	SEQ ID NO: 7	2797 bp
NOV2b, CG115907-04 DNA Sequence	GCCCCACAGTGAGAGGAAGGAAGGCAACAGTCGCCAGCAGCCGATGTGAAGACCGGACTCCGTCG CGCCCCCTCGCCGCTCTGCTGCGCCACATCGATGTTGTGTCCGCCGCTGCTCGCCCGGATCAC GATGAAGCCCCAAGGCTGTCCGTACCTGCAGCAAAGTTCTCGTCTGCTTTCACTGCTGGCC ATCCACCAGACTACTACTGCGGAAAGAATGGCATCGACATCTACAGCCTCACCGTGGACTCCA GGGTCTCATCCCGATTGCCCCACCGGTCTGTCACCGAGCCGAGTGGTCAATAGGGCCAATCTGT GCAGGAGGCCACCTTCCAGATGGAGCTGCCCAAGAAAGCCTTCATCACCACCTTCTCATGATC ATCGATGGCATGACCTACCCAGGATCATCAAGGAGAAGGCTGAAGCCAGGCACAGTACAGCG CAGCAGTGGCCAAAGGAAAGAGCGCTGGCCTCGTCAAGGCCACCGGGAGAAACATGGAGCAGTT CCAGGTGTCGGTCACTGTGGCTCCCAATGCCAAGATCACTTTGAGCTGGTCTATGAGGAGCTG CTCAAGCGGCGTTTGGGGGTGTACGAGCTGCTGCTGAAAGTGGGCCCCAGCAGCTGGTCAAGC ACCTGCAGATGGACATTACATCTTCGAGCCCCAGGGCATCAGCTTTCTGGAGACAGAGGCAC CTTCATGACCAACAGCTGGTAGACGCCCTCACCACCTGGCAGAATAAGACCAAGGCTCAATC CGGTTCAAGCAACACTTTCCAGCAGCAAAAGTCCCGAGAGCAGCAAGAAACAGTCTGGACG GCAACCTCATTATCCGCTATGATGTGGACCGGGCCATCTCCGGGGCTCCATTAGATCGAGAA CGGCTACTTTGTACACTACTTTGCCCGGAGGGCTAACCACAAATGCCCAAGAAATGTGGTCTTT GTCATTGACAAGAGCGGCTCCATGAGTGGCAGGAAATCCAGCAGACCCGGAAGCCCTAATCA AGATCCTGGATGACCTCAGCCCCAGAGACCAAGTTCAACCTCATCGTCTTCAGTACAGAAGCAAC TCAGTGGAGGCCATCACTGGTGCCAGCCTCAGCCGAGAACGTGAACAGGCCAGGAGCTTTGCT GCGGCATCCAGGCCCTGGGAGGGAACCAATCAATGATGCAATGCTGATGGCTGTGAGTTGC TGGACAGCAGCAACAGGAGGAGCGGCTGCCCAAGGGAGTGTCTCACTCATCTCTGCTCAC CGATGGCGACCCCACTGTGGGGGAGACTAACCCAGGAGCATCAGAATAACGTGCGGGAAGCT GTAAGTGGCGGTACAGCCTTTCTGCGTGGGCTTCGGTTTCAGACTCAGCTATGCTCTCTG AGAAGCTGGCACTGGCAATGGCGGCTTGGCCCGGCATCCATGAGGACTCAGACTCTGCCCT GCAGCTCCAGGACTTCTACAGGAAGTGGCAACCCACTGCTGACAGCAGTGACCTTCAGGTAC CCAAGCAATGCCGTGGAGGAGTCACTCAGAACAACTTCGGCTCCTCTCAAGGGCTCAGAGA TGGTGGTGGCTGGGAAGCTCCAGGACCGGGGCTGATGTGCTCACAGCCACAGTCACTGGGAA GCTGCCACACAGAACATCACTTTCAAACGGAGTCCAGTGTGGCAGAGCAGGAGGCGGAGTTC CAGAGCCCCAAGTATATCTTCCAACTTCATGGAGAGGCTCTGGGCATACCTGACTATCCAGC AGCTGCTGGAGCAAACTGTCTCCGATCCGATGCTGATCAGCAGGCCCTCCGGAACCAAGCGCT GAATTTATCACTTGCCTACAGCTTTGTACGCGCTCTCATCTATGGTAGTCACCAAAACCGAT GACCAAGAGCAGTCTCAAGTTGTGAGAAGCCCATGGAAGGCGAAAGTAGAAACAGGAATGTCC ACTCAGCTGGAGCTGCTGGCTCCCGATGAATTTAGACCTGGGTTCTCAGCTCCAGGCAACT TGGACTCCAGGACCTCTGATGTTCTGACCATGCTGCTTACCAACCCCTTCGCGCTGCTGGCC ATCTTGCCTGCTTACGACCACAGCCACCTCAAATCCTGATCCAGCTGTGTCTCGTGTATGA ATATGAGTATGAGAGGGAGAAGGCTGGGTTCTCATGATCGAAGTGACCTTCAAGAACCCCT GGTATGGGTTACGCATCCCTGAACAGTGGTGGTGAATCGGAACCGAAGAGCTCTGCGTAC AAGTGGAAAGGAGACCTATTCTCAGTGATGCCCGGCTGAAGATGACCATGGACAAGACGGGTC TCCTGCTGCTCAGTACCCAGACAAAGTGACCATCGGCTGTTGTTCTGGGATGGCCGTGGGA GGGGCTCCGCTCCTTCTGCTGACACTGACCGCTTCTCAGCCACGTTGGAGGGACCTTGGC CAGTTTTACAGGAGGTGCTCTGGGATCTCAGCAGCATCAGATGACCGCAGACGACGCTGA GGGTTACGGCAATGACCACTCTGCCACAGAGAGCGCAGGCTGGATTACAGGAGGGGCCCC GGGAGTGGAGATTTCTGCTGCTGTGGAGCTGTAGTTCTGATGGAAGGAGCTGTGCCCCCCC TGTACACTTGGCTTCCCCCTGCAACTGCAGGGCCGCTTCTGGGGCTGGACCACCATGGGAGG AAGAGTCCCACTCATTAATAAGAAAGGTGGTGTGAGCCTGA	

ORF Start: ATG at 130	ORF Stop: TAG at 2659
-----------------------	-----------------------

	SEQ ID NO: 8	843 aa	MW at 93770.6kD
NOV2b, CG115907-04 Protein Sequence	MKPPRPVRTCSKVLVLLSLAIHQTTAEKNGIDIYSLTVDSRVSSRFHTVVTSRVVRNANTV QEATFQMELEPKAFITNFSMIIDGMTYPGIIKEKAEQAQYSAAVAKGKSAGLVKATGRNMEQF QVSVSVAPNAKITFELVYEELLKRRLGVYELLKVRPQQLVKHLQMDIHIFEPQGISFLETEST FMTNQLVDALTTWQNKTKAHIRFKPTLSQQQKSPEQQTVDLGNLIIRYDVRDRAISGGSIQIEN GYFVHYFAPEGLTTMPKNVVFVIDKSGSMSGRKIQQTREALIKILDDLSPRDQFNLI VFSTEAT. QWRPSLVPSAENVNKARSFAAGIQALGGTNINDAMLMAVQLLDSSNQERLPEGSVSLIILLT DGDPTVGETNPRSIQNNVREAVSGRYSFLCLGFGFDVSYAFLEKLALDNGGLARRIHEDSDSAL QLQDFYQEVANPLLTAVTFEYPSNAVEEVTQNNFRLLPKGSEM VVAGKLQDRGPDVLTATVSGK LPTQNI TFQTESSVAEQEAEFQSPKYIFHNFMERLWAYLTIQQLLEQTVSASDADQQAALRNQAL NLSLAYSFVTPLTSMVVTKPDDQEQSQAEPKMEGESRNRNVHSAGAAGSRMNFPRGVLSSRQL GLPGPPDVPDHAAYHPFRRLAILPASAPPATSNPDPAVSRVMNQYEREKAGFSWLEVTFKNPL VWVHASPEHVVTNRNRSSAYKWKETLFSVMPGLKMTMDKTGLLLSDPKVTIIGLFWDGGRGE GLRLLLRDTRDFSSHVGGTLGQFYQEVWGSFAASDDGRRTL RVQGNDSATRERRLDYQEGFP GVEISCWSVEL		

5

	SEQ ID NO: 9	2914 bp	
NOV2c, CG115907-03 DNA Sequence	GCCCCACAGTGAGAGGAAGGAAGGCAACAGTCGCCAGCAGCCGATGTGAAGACCGGACTCCGTG CGCCCTCGCCGCTCTGCTGGCCACATCGATGTTGTGTCCGCGCCTGCTCGCCCGATCAC GATGAAGCCCCCAAGGCTGTCCGTACCTGCAGCAAGTTCTCGTCTGCTTTCAGTGTGGCC ATCCACCAGACTACTATGCGCGAAAGAATGGCATCGACATCTACAGCCTCACCGTGGACTCCA GGGTCTCATCCCGATTTGCCACACGGTCTGTCACAGCCGAGTGGTCAATAGGCGCAATACTGT GCAGGAGGCCACTTCCAGATGGAGCTGCCCAAGAAAGCCTTCATCAACCACTTCCATGATC ATCGATGGCATGACCTACCCAGGATCATCAAGGAGAAGGCTGAAGCCCAGGCACAGTACAGCG CAGCAGTGGCCAAGGGAAGAGCGCTGGCCTCGTCAAGGCCACCGGGAGAAACATGGAGCAGTT CCAGGTGTCTGCTCAGTGTGGCTCCCAATGCCAAGATCACTTTGAGCTGGTCTATGAGGAGCTG CTCAAGCGCGCTTTGGGGGTGTACGAGCTGCTGCTGAAAGTGCGGCCCCAGCATGCTGGTCAAGC ACCTGCAGATGGACATTACATCTTCGAGCCCCAGGGCATCAGCTTCTGAGACAGAGAGCAC CTTTCATGACCAACAGCTGGTAGACGCCCTCACCACCTGGCAGAATAAGACCAAGGCTCACATC CGGTTCAAGCCAACACTTTCCAGCAGCAAAAGTCCCAAGAGCAGCAAGAAACAGTCTGGACG GCAACCTCATTATCCGCTATGATGTGGACCGGGCCATCTCCGGGGCTCCATTAGATCGAGAA CGGCTACTTTGTACACTACTTTGCCCCGAGGGCCTAACCACAATGCCAAGAAATGTGGTCTTT GTCATTGACAAGAGCGGCTCCATGAGTGGCAGGAAAATCCAGCAGACCCGGGAAGCCCTAATCA AGATCCTGGATGACCTCAGCCCCAGAGACCAAGTTCAACCTCATGCTCTTCAGTACAGAAGCAAC TCAGTGGAGGCCATCAGTGGTGGCAGCCTCAGCCGAGAACTGTAACAAGGCCAGGAGCTTTGCT GCGGGCATCCAGGCCCTGGGAGGGACCAATCAATGATGCAATGTCTGATGGCTGTGAGTTGC TGGACAGCAGCAACAGGAGGAGCGGCTGCCGAAGGGAGTGTCTCATCTCCTGCTCAGC CGATGGCGACCCCACTGTGGGGGAGACTAAGCCAGGAGCATCCAGAATAAGTCCGGGAAGCT GTAAGTGGCCGCTACAGCTCTTCTGCTGGCTTCCGTTTCAGCTGAGCTATGCTCTCCTGG AGAAGCTGGCACTGGACAATGGCGGCTGGCCCGCGCATCCATGAGGACTCAGACTCTGCCCT GCAGCTCCAGGACTTCTACCAAGGAAGTGGCCAAACCACTGCTGACAGCAGTGAACCTCGAGTAC CCAAGCAATGCCGTGGAGGAGTCACTCAGAACAACTTCCGGCTCTCTTCAAGGGCTCAGAGA TGGTGGTGGCTGGGAAGCTCCAGGACCGGGGGCTGATGTGCTCAGGCCACAGTCACTGGGAA GCTGCCACACAGAACATCACTTCCAAACGAGTCCAGTGTGGCAGAGCAGGAGCGGAGTTT CAGAGCCCCAAGTATATCTTCCCAACTTCATGGAGAGGCTCTGGGCATACCTGACTATCCAGC AGCTGCTGGAGCAAACTGTCTCCGCATCCGATGCTGATCAGCAGGCCCTCCGGAACCAAGCGCT GAATTTATCACTTGCCCTACAGCTTTGTACCGCTCTCACATCTATGGTAGTACCAAACCGAT GACCAAGAGCAGTCTCAAGTTGCTGAGAAGCCCATGGAAGGCGAAAGTAGAAACAGGAATGTCC ACTCAGCTGGAGCTGTGGCTCCCGGATGAATTTACAGCTGGGGTTCTCAGCTCCAGGCACT TGGACTCCCAGGACCTCTGATGTTCTGACCATGTGCTTACACCCCTTCCGCGCTCTGGCC ATCTGCTGCTTTCAGCAACACAGCCACCTCAAATCCTGATCCAGCTGTGTCTCGTGTCTATGA ATATGTCTGCCATCCTGCCACTGCTGGGAGAGTGTGGAGCGGCTCTGTGTGGACCCCAACA CCGCCAGGGGCCAGTGAACCTGCTCTCAGACCTGAGCAAGGGGTTGAGGTGACTGGCCAGTAT GAGAGGGAGAAGGCTGGGTTCTCATGGATCGAAGTGACCTTCAAGAACCCCTGATATGGGTTT ACGCATCCCCTGAACACGTTGGTGGTGAAGTCCGGAACCGAAGAAGCTCTGCGTCAAGTGAAGGA GACGCTATTCTCAGTGATGCCCGGCTGAAGATGACCATGGACAAGGAGTCTCTGCTGCTGCTC AGTGACCCAGACAAAGTGACCATCGGCCCTGTTGTTCTGGGATGGCCGTGGGGAGGGGCTCCGGC TCCTTCTGCTGACACTGACCGCTTCTCAGCCACGTTGGAGGGACCTTGGCCAGTTTTACCA GGAGGTGCTCTGGGGATCTCAGCAGCATCAGATGACGGCAGACGACGCTGAGGGTTTACGGGC AATGACCACTCTGCCACAGAGAGCGCAGGCTGGATTACAGGAGGGGCCCCCGGAGTGGAGA		

TTTCCTGCTGGTCTGTGGAGCTGTAGTTCTGATGGAAGGAGCTGTGCCACCCCTGTACACTTGG CTTCCCCCTGCAACTGCAGGGCCGCTTCTGGGGCCTGGACCACCATGGGGAGGAAGAGTCCCAC TCATTACAAATAAGAAAGGTGGTGTAGCCTGA		
ORF Start: ATG at 130		ORF Stop: TAG at 2776

	SEQ ID NO: 10	882 aa	MW at 97921.2kD
NOV2c, CG115907-03 Protein Sequence	MKPPRPVRTCCKVLVLLSLLAHQTTTAEKNGIDIYSLTVDSRVSSRFHTVVTSRVVNRANTV QEATFQMELEPKKAFITNFSMIIDGMTYPGIKEKAEAAQYSAAVAKGKSAGLVKATGRNMEQF QVSVSVAPNAKIFTELVEELLKRRLLGVYELLKVRPQQLVKHLOMDIHFEPQGISFLETEST FMNQLVDALTTWQNKTKAHIRFKPTLSQQQKSPEQOETVLDGNLIIRYDVDRALSGGSIQIEN GYFVHYFAPEGLTTPKNNVVFIDKSGSMSGRKIQQTREALIKILDDLSPRDQPNLIVFSTEAT QWRPSLVPASAENVNKARSFAAGIQALGGTININDAMLMVQLLDSSNQEERLPEGSVSLIILLT DGDPTVGETNPRSIQNNVREAVSGRYSFLCLGFGFDVSYAFLEKLALDNGGLARRIHEDSDSAL QLQDFYQEVANPLLTAVTFEYPSNAVEEVTQNNFRLLFKGSEMVVAGKLQDRGPDVLTATVSGK LPTQNI TFQTESSVAEQEAEFQSPKYYI FHNFMERLWAYLTIQQLLEQTVSASDADQALRNQAL NLSLAYSFVTPLTSMVVTKPDDEQSQVAEKPMEGESRNRNVHSAGAAGSRMFRPGVLSSRQL GLPGPPDVPDHAAYHPFRRLAILPASATPATSNPDPAVSRVMMSAILPLPGQSVRLCVDPRH RQGPVNLLSDPEQGEVETGQYEREKAGFSWIEVTFKNPLVWVHASPEHVVTNRNRSSAYKWKE TLFSVMPGLKMTMDKTGLLLSLDPPDKVTIGLLFWDGRGEGLRLLRLRDTDRPSSHVGGTGLQGFYQ EVLWGSPAASDDGRRTLRLVQGNHSATRERRLDYQEGPPGVEISCWSVEL		

5

	SEQ ID NO: 11	2968 bp
NOV2d, CG115907-02 DNA Sequence	GCCCCACAGTGAGAGGAAGGAAGGCAACAGTCGCCAGCAGCCGATGTGAAGACCGGACTCCGTG CGCCCTCGCCGCTCTGCCTGGCCACATCGATGTTGTGTCCGCCGCTGCTCGCCCGGATCAC GATGAAGCCCCAAGGCTGTCCGTACCTGCAGCAAAGTTCTGCTCTGCTTTCACTGCTGGCC ATCCACAGACTACTACTGCCGAAAGAAATGGCATCGACATCTACAGCCTCACCGTGGACTCCA GGTCTCATCCCGATTGCCCCACCGTCTGCACAGCCGAGTGGTCAATAGGGCCAATACTGT GCAGGAGGCCACCTTCCAGATGGAGCTGCCCAAGAAAGCCTTCATACCAACTTCTCCATGATC ATCGATGGCATGACCTACCCAGGGATCATCAAGGAGAAGGCTGAAGCCAGGCACAGTACAGCG CAGCAGTGGCCAAGGGAAAGAGCGCTGCCCTCGTCAAGGCCACCGGAGAAACATGGAGCAGTT CCAGGTGTCCGTGAGTGTGGCTCCCAATGCCAAGATCACCTTTGAGCTGGTCTATGAGGAGCTG CTCAAGCGGCGTTTGGGGGTGTACGAGCTGCTGCTGAAGTGGCGGCCAGAGCTGGTCAAGC ACCTGCAGATGGACATTACATCTTCGAGCCCCAGGGCATCAGCTTTCTGGAGACAGAGAGCAC CTTCATGACCAACAGCTGGTAGACGCCCTCACCACTGGCAGAAATAAGACCAAGGCTCACATC CGGTTCAAGCCAACACTTTCCAGCAGCAAAGTCCCCAGAGCAGCAAGAAACAGTCTTGGACG GCAACCTCATTATCCGCTATGATGTGGACCGGGCCATCTCCGGGGCTCCATTGATCGAGAA CGGCTACTTTGTGACTACTTTGCCCCGAGGGGCTAACCAATGCCAAGAAATGTGGTCTTT GTCATTGACAAGAGCGGCTCCATGAGTGGCAGGAAAATCCAGCAGACCCGGGAAGCCCTAATCA AGATCCTGGATGACCTCAGCCCCAGAGACCAAGTTCAACCTCATCGTCTTCAGTACAGAACCAAC TCAGTGAGGCCATCACTGGTCCAGCCTCAGCCGAGAACGTGAACAGGCCAGGAGCTTTGTCT GCGGGCATCCAGGCCCTGGGAGGGACCAACATCAATGATGCAATGCTGATGGCTGTGATGGTGC TGGACAGCAGCAACAGGAGGAGCGGCTGCCGAAGGGAGTGTCTCACTCATCATCTGCTCAC CGATGGCGACCCACTGTGGGGGAGACTAACCCAGGAGCATCCAGAATAACGTGCGGGAAGCT GTAAGTGGCCGGTACAGCCTCTTCTGCTGGGCTTCGGTTTCAGCTCAGCTATGCCCTTCTGG AGAAGCTGGCACTGGACAATGGCGGCTGGCCCCGGCGCATCCATGAGGACTCAGACTGTGCCCT GCAGCTCCAGGACTTCTACAGGAAGTGGCCAAACCACTGCTGACAGCAGTGACCTTCGAGTAC CCAAGCAATGCCGTGGAGGAGGTCACTCAGAACAACTTCGGGCTCCTTCAAGGGCTCAGAGA TGGTGGTGGCTGGGAAGCTCCAGGACCGGGGGCTGATGTGCTCAGGCCACAGTCACTGAGTGGAA GCTGCCTACACAGAACATCACTTTCCAAACGGAGTCCAGTGTGGCAGAGCAGGAGGCGAGTTT CAGAGCCCCAAGTATATCTTCCAACTTTCATGGAGAGGCTCTGGGCATACCTGATATCCAGC AGCTGCTGGAGCAAAGTGTCTCCGATCCGATGCTGATCAGCAGGCCCTCCGGAACCAAGCGCT GAATTTATCACTTGCTACAGCTTTGTACAGCCTCTCACATCTATGCTAGTACGTAACCAACCGAT GACCAAGAGCAGTCTCAAGTTGCTGAGAAGCCCATGGAAGGCGAAAGTAGAAACAGGAATGTCC ACTCAGCTGGAGCTGCTGGCTCCCGGATGAATTCAGACCTGGGGTCTCAGCTCCAGGCAACT TGGACTCCAGGACCTCCTGATGTTCTGACCATGCTGCTTACCAACCCCTTCCGCGTCTGGCC ATCTTGCTGCTTCCAGCAACCAAGCCACCTCAATCCTGATCCAGCTGTGTCTCGTGTCTATGCA ATATGAAAATCGAAGAAACACCATGACAAACCAACCCAGCCCCATACAGGCTCCCTCTGC CATCTGCCACTGCTGGGCGAGTGTGGAGCGGCTCTGTGTGGACCCAGACACCGCCAGGGG CCAGTGAACCTGCTCTCAGACCTGAGCAAGGGGTTGAGGTGACTGGCCAGTATGAGAGGGAGA AGGCTGGGTTCTCATGGATCGAAGTACCTTCAAGAACCCCTGTTATGGGTTACGCATCCCC TGAACACGTGGTGTGACTCGGAACCGAAGAGCTCTGCGTACAAGTGAAGGAGACGCTATTTC TCAGTGATGCCGGGCTGAAGATGACCATGGACAAGACGGGTCTCTGCTGCTCAGTGACCCAG	

	ACAAAGTGACCATCGGCCTGTTGTTCTGGGATGGCCGTGGGGAGGGGCTCCGGCTCCTTCTGCG TGACACTGACCGCTTCTCCAGCCACGTTGGAGGGACCTTGGCCAGTTTTACCAGGAGGTGCTC TGGGGATCTCCAGCAGCATCAGATGACGGCAGACGCACGCTGAGGGTTTCAGGGCAATGACCACT CTGCCACCAGAGAGCGCAGGCTGGATTACCAGGAGGGCCCCCGGAGTGGAGATTTCTTGCTG GTCTGTGGAGCTGTAGTTCTGATGGAAGGAGCTGTGCCACCCTGTACACTTGGCTTCCCCCTG CAACTGCAGGGCCGCTTCTGGGGCTGGACCACCATGGGGAGGAAGAGTCCCACTCATTACAAA TAAAGAAAGGTGGTGTGAGCCTGA	
	ORF Start: ATG at 130	ORF Stop: TAG at 2830

	SEQ ID NO: 12	900 aa	MW at 99856.4kD
NOV2d, CG115907-02 Protein Sequence	MKPPRPVRTCSKVLVLLSLAIHQTTAEKNGIDIYSLTVDSRVSSRFHTVTVSRVNRANTV QEATFQMELEPKKAFITNFSMIIDGMTYPGIIKEKAEQAQYSAAVAKGKSAGLVKATGRNMEQF QVSVSVAPNAKTFELVYEELLKRLGVYELLLKVRPQQLVKHLQMDIHIFEPQGISFLETEST FMINQLVDALT'TWQNKTKAHIRFKPTLSQQQKSPEQQETVLDGNLIIRYDVEDRAISGGSIQIEN GYFVHYFAPEGLTTPKKNVVFVIDKSGSMSGRKIQQTREALIKILDDLSPRDQFNLIIVFSTEAT QWRPSLVPASAENVNKARSPAAGIQALGGTNINDAMLMVQLDSSNQEERLPEGSVSLIILLT DGDPTVGETNPRSIQNNVREAVSGRYSLFCLGFGFDVSYAFLEKLALDNGGLARRIHEDSDSAL QLQDFYQEVANPLLTAFTFEYPSNAVEEVTQNNFRLLFKGSEMVMVAGKLQDRGPDVLTATVSGK LPTQNTFQTESSVAEQEAEFQSPKIFHNFMERLWAYLTIQQLLEQTVSASDADQQALRNQAL NLSLAYSFVTPLTSMVVTKPDDEQSQVAEKPMEGESRNRNVHSAGAAGSRMFRPGVLSSRQL GLPGPPDVPDHAAHYHFRRLAILPASAPPATSNPDPAVSRVMNMKIEETMTTQTPTAPIQAPSA ILPLPGQSVERLCVDFRHRQGFVNLLSDPEQGVETGQYEREKAGFSWIEVTFKNPLVWVHASP EHVVTNRNRSSAYKWKETLFSVMPGLKMTMDKTGLLLSDPDKVTIGLLFWDGRGEGRLRLLR DTRFSSHVGGTLGQFYQEVWLGSPAASDDGRRTLVRQGNHDSATRERRLDYQEGPPGVEISCW SVEL		

- 5 Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 2B.

Table 2B. Comparison of NOV2a against NOV2b through NOV2d.		
Protein Sequence	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV2b	1..930 1..843	804/930 (86%) 813/930 (86%)
NOV2c	1..930 1..882	881/930 (94%) 881/930 (94%)
NOV2d	1..930 1..900	900/930 (96%) 900/930 (96%)

- 10 Further analysis of the NOV2a protein yielded the following properties shown in Table 2C.

Table 2C. Protein Sequence Properties NOV2a	
SignalP analysis:	Cleavage site between residues 29 and 30

PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 8; pos.chg 3; neg.chg 0  H-region: length 3; peak value 3.04  PSG score: -1.36</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): 0.51  possible cleavage site: between 27 and 28</p> <p>&gt;&gt;&gt; Seems to have no N-terminal signal peptide</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 1  Tentative number of TMS(s) for the threshold 0.5:  1  Number of TMS(s) for threshold 0.5: 0  PERIPHERAL Likelihood = 2.01 (at 578)  ALOM score: -0.64 (number of TMSs: 0)</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 2 Hyd Moment(75): 3.74  Hyd Moment(95): 8.90 G content: 0  D/E content: 1 S/T content: 6  Score: -1.36</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  R-2 motif at 18 VRT CS</p> <p>NUCDISC: discrimination of nuclear localization signals  pat4: none  pat7: none  bipartite: none  content of basic residues: 10.4%  NLS Score: -0.47</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals:  XXRR-like motif in the N-terminus: KPPR</p> <p>none</p> <p>SKL: peroxisomal targeting signal in the C-terminus:  none</p> <p>PTS2: 2nd peroxisomal targeting signal: none</p> <p>VAC: possible vacuolar targeting motif: none</p> <p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif:  type 1: none  type 2: none</p> <p>NMYR: N-myristoylation pattern : none</p> <p>Prenylation motif: none</p>
--------------------	--

	<p>memYQRL: transport motif from cell surface to Golgi: none</p> <p>Tyrosines in the tail: none</p> <p>Dileucine motif in the tail: none</p> <p>checking 63 PROSITE DNA binding motifs: none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p> <p>checking 33 PROSITE prokaryotic DNA binding motifs: none</p> <p>NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination Indication: cytoplasmic Reliability: 70.6</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions total: 0 residues</p> <p>-----</p> <p>Final Results (k = 9/23):</p> <p>60.9 %: mitochondrial 8.7 %: cytoplasmic 8.7 %: extracellular, including cell wall 8.7 %: peroxisomal 4.3 %: vacuolar 4.3 %: Golgi 4.3 %: nuclear</p> <p>&gt;&gt; indication for CG115907-01 is mit (k=23)</p>
--	--

- A search of the NOV2a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded
- 5 several homologous proteins shown in Table 2D.

<b>Table 2D. Geneseq Results for NOV2a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV2a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
ABB09709	Amino acid sequence of a human PK-120 polypeptide - Homo sapiens, 930 aa. [WO200212495-A1, 14-FEB-2002]	1..930 1..930	930/930 (100%) 930/930 (100%)	0.0
ABB09708	Sequence of H4P heavy chain of inter alpha trypsin inhibitor - Homo sapiens, 930 aa. [WO200212495-A1, 14-FEB-2002]	1..930 1..930	928/930 (99%) 929/930 (99%)	0.0
ABB09711	Sequence of H4P heavy chain of inter-alpha-inhibitor protein - Sos sp, 921 aa. [WO200212495-A1, 14-FEB-2002]	13..930 12..921	663/924 (71%) 758/924 (81%)	0.0
ABB09707	Sequence of H4P heavy chain of inter-alpha-inhibitor protein - Rattus sp, 933 aa. [WO200212495-A1, 14-FEB-2002]	1..930 1..933	615/941 (65%) 728/941 (77%)	0.0
ABB09706	Sequence of H4P heavy chain of inter-alpha-inhibitor protein - Rattus sp, 932 aa. [WO200212495-A1, 14-FEB-2002]	1..930 1..932	600/941 (63%) 715/941 (75%)	0.0

In a BLAST search of public sequence databases, the NOV2a protein was found to have homology to the proteins shown in the BLASTP data in Table 2E.

5

<b>Table 2E. Public BLASTP Results for NOV2a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV2a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q14624	Inter-alpha-trypsin inhibitor heavy chain H4 precursor (ITI heavy chain H4) (Inter-alpha-inhibitor heavy chain 4) (Inter-alpha-trypsin inhibitor family heavy chain-related protein) (IHRP) (Plasma kallikrein sensitive glycoprotein 120) (PK-120) (GP120) (PRO1851) [Contains: GP57] - Homo sapiens (Human), 930 aa.	1..930 1..930	929/930 (99%) 929/930 (99%)	0.0

JX0368	inter-alpha-trypsin inhibitor heavy chain-related protein precursor - human, 930 aa.	1..930 1..930	928/930 (99%) 929/930 (99%)	0.0
P79263	Inter-alpha-trypsin inhibitor heavy chain H4 precursor (ITI heavy chain H4) (Inter-alpha-inhibitor heavy chain 4) (Inter-alpha-trypsin inhibitor family heavy chain-related protein) (IHRP) (Major acute phase protein) (MAP) - Sus scrofa (Pig), 921 aa.	13..930 12..921	663/924 (71%) 758/924 (81%)	0.0
Q91W60	Inter alpha-trypsin inhibitor, heavy chain 4 - Mus musculus (Mouse), 941 aa.	1..930 1..941	625/958 (65%) 743/958 (77%)	0.0
O54882	PK-120 - Mus musculus (Mouse), 942 aa.	1..930 1..942	621/957 (64%) 740/957 (76%)	0.0

PFam analysis indicates that the NOV2a protein contains the domains shown in the Table 2F.

Table 2F. Domain Analysis of NOV2a			
Pfam Domain	NOV2a Match Region	Identities/ Similarities for the Matched Region	Expect Value
vwa	274..457	34/209 (16%) 125/209 (60%)	1.1e-08

5

### Example 3.

The NOV3 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 3A.

Table 3A. NOV3 Sequence Analysis			
	SEQ ID NO: 13	1365 bp	
NOV3a, CG139008-01 DNA Sequence	ATGCTGCGGATCCTGTGCCTGGCACTCTGCAGCCTGCTGACTGGCAGCGAGCTGACCCTGGGG CACTGCTGCGGTGGGCATGGACATCATGAACCGTGAGGTCCAGAGCGCCATGGATGAGAGTCA TATCCTGGAGAAGATGGCAGCCGAGGCAGGCAAGAAACAGCCAGGGATGAAACCTATCAAGGGC ATCACC AATTTGAAGGTGAAGGATGTCCAGCTGCCCCTCATCACTGAACTTTGTACCTGGAG TGGGCATCTTCCAATGTGTGTCCACAGGCATGACCGTCACTGGCAAGAGCTTCATGGGAGGGAA CATGGAGATCATCGTGGCCCTGAACATCACAGCCACCAACCGGCTTCTGCGGGATGAGGAGACA GGCCTCCCCGTGTTCAAGAGTGAGGGCTGTGAGGTCATCCTGGTCAATGTGAAGACTAACCTGC CTAGCAACATGCTCCCCAAGATGGTCAACAAGTTCTGGACAGCACCCCTGCACAAAGTCTCCC TGGGCTGATGTGTCCCGCCATCGATGCAGTCCTGGTGTATGTGAACAGGAAGTGGACCAACCTC AGTGACCCCATGCCTGTGGGCCAGATGGGCACCGTCAAATATGTTCTGATGTCCGCACCAAGCCA CCACAGCCAGCTACATCCAACCTGGACTTCAGTCTGTGGTGCAGCAGCAAAAGGGCAAAACCAT CAAGCTTGCTGATGCCGGGGAGGCCCTCACGTTCCCTGAGGGTTATGCCAAAGGCTCGTCGAG		

	CTGCTGCTCCAGCCACCTTCCTCTCTGCAGAGCTTGCCCTTCTGCAGAAGTCCTTTTCATGTGA ATATCCAGGATACAATGATTGGTGAGCTGCCCCACAAACCACCAAGACCTGGCTCGCTTCAT TCCTGAAGTGGCTGTAGCTTATCCCAAGTCAAAGCCCTTGACGACCCAGATCAAGATAAAGAAG CCTCCCAAGGTCATATGAAGACAGGCAAGAGCCTGCTGCACCTCCACAGCACCTGGAGATGT TCGCAGCTCGGTGGCGGAGCAAGGCTCCAATGTCCCTCTTCTCCTAGAAGTGCACCTCAATCT GAAGGTCCAGTACTCAGTGCATGAGAACCAGCTGCAGATGGCCACTTCTTTGGACAGATTACTG AGCTTGTCCCGAAGTCTCATCGATTGGCACTTCAATGAGAGGGAATTAAGTGGCTTCATCA CCAGCTATCTCGAAGAAGCCTACATCCAGTTGTCAATGATGTGCTTCAAGTGGGGCTCCACT CCCGGACTTCTGGCCATGAATTACAACCTGGCTGAGCTGGACATAGTAGAGCTTGGGGGCATC ATGGAACCTGCCGACATATGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 1363

	SEQ ID NO: 14	454 aa	MW at 49801.1kd
NOV3a, CG139008-01 Protein Sequence	MLRILCLALCSLLTGTRADPGALLRLGMDIMNREVQSAMDESHILEKMAAEAGKKQPGMKPIKG ITNLKVKDVQLPVIITLNFVPGVGIFQCVSTGMTVTGKSFMGNNMEIIVALNITATNRLLRDEET GLPVFKSEGCEVILVNVKTNLPSNMLPKMVNKFLLDSTLHKVLPGLMCPAIDAVLVVYVNRKWTNL SDPMPVQMGTVKYVILMSAPATTASYIQLDSPVQQQKGTIKLADAGEALTFPEGYAKGSSQ LLLPLATFLSAELALLQKSFHVNIQDTMIGELPPQTTKTLARFIPEVAVAYPKSKPLTTQIKIKK PPKVTMTKGKSLHLHSTLEMFARWRSKAPMSLFLLLEVHFNLKVQYSVHENQLQMATSLDRLL SLSRKSSSIGNFNERELTGFITSYLEEAYIPVVNDVLQVGLPLPDFLAMNYNLAELDIVELGGI MEPADII		

5

	SEQ ID NO: 15	1374 bp	
NOV3b, 233028732 DNA Sequence	AGATCTATGCTGCGGATCCTGTGCTGGCACTCTGCAGCCTGCTGACTGGCACGCGAGCTGACC CTGGGGCACTGCTGCGGTTGGGCATGGACATCATGAACCGTGAGGTCAGAGCGCCATGGATGA GAGTCATATCCTGGAGAAGATGGCAGCGAGGCAAGAAACAGCCAGGGATGAAACCTATC AAGGGCATCACCATTGAAGGTGAAGGATGTCCAGCTGCCGTCATCACTGAACCTTTGTAC CTGGAGTGGGCATCTTCCAATGTGTGTCCACAGGCATGACCGTCACTGGCAAGAGCTTCATGGG AGGGAAATGGAGATCATCGTGGCCCTGAACATCACAGCCACCAACCGCTTCTGCGGGATGAG GAGACAGGCCCTCCCGTGTCAAGAGTGAGGGCTGTGAGGTCACTCGGTCAATGTGAAGACTA ACCTGCTAGCAACATGCTCCCAAGATGGTCAACAAGTTCCCTGGA CAGCACCCCTGCACAAAGT CCTCCCTGGGCTGATGTGTCGCCCATCGATGCAGTCTGCTGTATGTGAACAGGAAGTGGACC AACCTCAGTGACCCCATGCTGTGGGCCAGATGGGCACCGTCAAATATGTTCTGATGTCCGCAC CAGCCACCACAGCCAGCTACATCCAACCTGGACTTCAGTCTGTGGTGCAGCAGCAAAAGGGCAA AACCATCAAGCTTGCTGATGCCGGGGAGGCCCTCAGTTCCTGAGGGTTATGCCAAAGGCTCG TCGCAGCTGCTGCTCCAGCCACCTTCCTCTCTGCAGAGCTTGCCCTTCTGCAGAAGTCCTTTC ATGTGAATATCCAGGATACAATGATTGGTGAGCTGCCCCACAAACCACCAAGACCTGGCTCG CTTCATTCTGAAAGTGGCTGTAGCTTATCCCAAGTCAAAGCCCTTGACGACCCAGATCAAGATA AAGAAGCTCCCAAGGTCATATGAAGACAGGCAAGAGCCTGCTGCACCTCCACAGCACCTGG AGATGTTGCGAGCTCGGTGGCGGAGCAAGGCTCCAATGTCCCTCTTCTCCTAGAAGTGCACCT CAATCTGAAGGTCCAGTACTCAGTGCATGAGAACCAGCTGCAGATGGCCACTTCTTTGGACAGA TTACTGAGCTTGTCCCGAAGTCTCATCGATTGGCACTTCAATGAGAGGGAATTAAGTGGCT TCATCACCAGCTATCTCGAAGAAGCCTACATCCAGTTGTCAATGATGTGCTTCAAGTGGGGCT CCCACTCCCGGACTTCTGGCCATGAATTACAACCTGGCTGAGCTGGACATAGTAGAGCTTGGG GGCATCATGGAACCTGCCGACATACCTCGAG		
	ORF Start: at 1		ORF Stop: end of sequence

	SEQ ID NO: 16	458 aa	MW at 50286.7kd
NOV3b, 233028732 Protein Sequence	RSMRLILCLALCSLLTGTRADPGALLRLGMDIMNREVQSAMDESHILEKMAAEAGKKQPGMKPI KGITNLKVKDVQLPVIITLNFVPGVGIFQCVSTGMTVTGKSFMGNNMEIIVALNITATNRLLRDE ETGLPVFKSEGCEVILVNVKTNLPSNMLPKMVNKFLLDSTLHKVLPGLMCPAIDAVLVVYVNRKWT NLSDPMPVQMGTVKYVILMSAPATTASYIQLDSPVQQQKGTIKLADAGEALTFPEGYAKGS SQLLLPATFLSAELALLQKSFHVNIQDTMIGELPPQTTKTLARFIPEVAVAYPKSKPLTTQIKIK KKPPKVTMTKGKSLHLHSTLEMFARWRSKAPMSLFLLLEVHFNLKVQYSVHENQLQMATSLDR LLSLSRKSSSIGNFNERELTGFITSYLEEAYIPVVNDVLQVGLPLPDFLAMNYNLAELDIVELG GIMEPADILE		

10

	SEQ ID NO: 17	1226 bp	
NOV3c, CG139008-02 DNA Sequence	ATGCTGCGGATCCTGTGCCTGGCACTCTGCAGCCTGCTGACTGGCACGCGAGCTGACCCCTGGGG CACTGCTGCGGTTGGGCATGGACATCATGAACCGTGAGGTCCAGAGCGCCATGGATGAGAGTCA TATCCTGGAGAAGATGGCAGCCGAGGCAAGAAAAGCCAGGGATGAAACCTATCAAGGGC ATACCAATTGTAAGGTGAAGGATGTCCAGCTGCCCGTCATCACACTGAACTTTGTACCTGGAG TGGGCATCTTCCAATGTGTGTCCACAGGCATGACCGTCACTGGCAAGAGCTTCATGGGAGGGAA CATGGAGATCATCGTGGCCCTGAACATCAAGCCACCAACCGGCTTCTGCGGGATGAGGAGACA GGCTCCCCGTGTTCAAGAGTGAGGGCTGTGAGGTTCATCCTGGTCAATGTGAAGACTAACCTGC CTAGCAACATGCTCCCCAAGATGGTCAACAAGTTCCTGGACAGCACCCCTGCACAAAGTCTCCCT TGGGCTGATGTGTCCCGCCATCGATGCAGTCTCTGGTGTATGTGAACAGGAAGTGGACCAACCTC AGTGACCCCATGCTGTGGGCCAGATGGGCACCGTCAAATATGTTCTGATGTCCGCACCAAGCCA CCACAGCCAGCTACATCCAACCTGGACTTCAGTCTGTGGTGCAGCAGCAAAAGGGCAAAACCAT CAAGCTTGCTGATGCGGGGAGGCCCTCACGTTCCCTGAGGGTTATGCCAAAGGCTCGTCCGAG CTGCTGCTCCAGCCACCTTCTCTCTGTCAGAGCTTGCCCTTCTGCAGAAGTCTTTTCATGTGA ATATCCAGGATACAATGATTGGTGAGCTGCCCCCAAAACCAAGACCCCTGGCTCGCTTCAT TCCTGAAGTGGCTGTAGCTTATCCCAAGTCAAAGCCCTTGACGACCCAGATCAAGATAAAGAAG CCTCCCAAGGTCACTATGAAGACAGGCAGAGCCCTGCTGCACCTCCACAGCACCCCTGGAGATGT TCGCAGCTCGGTGGCGGAGCAAGGCTCCAATGTCCCTCTTTCTCTAGAAGTGCACCTTCAATCT GAAGTCCAGTACTCAGTGCATGAGAACCAGCTGCAGATGGCCACTTCTTTGGACAGGAGAGGG AATTAAGTGGCTTCATCACCAGTATCTCGAAGAGCCTACATCCAGTGTGTCAATGATGTGCTT TCAGTGGGCT		
	ORF Start: ATG at 1		ORF Stop: TAA at 1156

	SEQ ID NO: 18	385 aa	MW at 42216.5kD
NOV3c, CG139008-02 Protein Sequence	MLRLCLALCSLLTGTRADPGALLRLGMDIMNREVQSAMDESHILEKMAEAGKKQPGMKPIKG ITNLKVKDVQLPVITLNFVPGVGIFQCVSTGMTVTGKSFMGGNMEIIVALNITATNRLRDEET GLPVFKSEGCEVILVNVKTNLPSNMLPKMVNKFLLDSTLHKVLPGLMCPAIDAVLVYVNRKWTNL SDPMPVQMGTVKYVLMSAPATTASYIQDLDFSPVVQQQKGKTIKLADAGEALTPEGYAKGSSQ LLLPATFLSAELALLQKSFHVNIQDTMIGELPPQTTKTLARFIPEVAVAYPKSKPLTQIKIKK PPKVTMKTGKSLHLHSTLEMFARWRSKAPMSLFLLEVHFNLKQVYSVHENQLQMATSLDRRG N		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 3B.

Table 3B. Comparison of NOV3a against NOV3b and NOV3c.		
Protein Sequence	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV3b	1..454 3..456	454/454 (100%) 454/454 (100%)
NOV3c	1..382 1..382	382/382 (100%) 382/382 (100%)

10

Further analysis of the NOV3a protein yielded the following properties shown in Table 3C.

Table 3C. Protein Sequence Properties NOV3a	
SignalP analysis:	Cleavage site between residues 19 and 20
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 3; pos.chg 1; neg.chg 0  H-region: length 13; peak value 9.26  PSG score: 4.86</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): 4.69  possible cleavage site: between 18 and 19</p> <p>&gt;&gt;&gt; Seems to have a cleavable signal peptide (1 to 18)</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 19  Tentative number of TMS(s) for the threshold 0.5:  2  Number of TMS(s) for threshold 0.5: 1  INTEGRAL Likelihood = -2.87 Transmembrane  169 - 185  PERIPHERAL Likelihood = 1.59 (at 255)  ALOM score: -2.87 (number of TMSs: 1)</p> <p>MTOP: Prediction of membrane topology (Hartmann et al.)  Center position for calculation: 9  Charge difference: -2.0 C( 0.0) - N( 2.0)  N &gt;= C: N-terminal side will be inside</p> <p>&gt;&gt;&gt; membrane topology: type 1a (cytoplasmic tail 186 to 454)</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 2 Hyd Moment(75): 7.85  Hyd Moment(95): 8.62 G content: 1  D/E content: 1 S/T content: 3  Score: -2.21</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  R-2 motif at 27 TRA DP</p> <p>NUCDISC: discrimination of nuclear localization signals  pat4: none  pat7: none  bipartite: none  content of basic residues: 9.9%  NLS Score: -0.47</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals:  XXRR-like motif in the N-terminus: LRIL</p> <p>none</p> <p>SKL: peroxisomal targeting signal in the C-terminus:  none</p>

	<p>PTS2: 2nd peroxisomal targeting signal: none</p> <p>VAC: possible vacuolar targeting motif: none</p> <p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif:            type 1: none            type 2: none</p> <p>NMYR: N-myristoylation pattern : none</p> <p>Prenylation motif: none</p> <p>memYQRL: transport motif from cell surface to Golgi:  none</p> <p>Tyrosines in the tail: too long tail</p> <p>Dileucine motif in the tail: found            LL at 257            LL at 258            LL at 270            LL at 332            LL at 356</p> <p>checking 63 PROSITE DNA binding motifs: none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p> <p>checking 33 PROSITE prokaryotic DNA binding motifs:  none</p> <p>NNCN: Reinhardt's method for Cytoplasmic/Nuclear  discrimination            Indication: cytoplasmic            Reliability: 89</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions  27 M 0.73  28 D 0.73  29 I 0.73  30 M 0.73  31 N 0.73  32 R 0.73  33 E 0.73  34 V 0.73  35 Q 0.73  36 S 0.73  37 A 0.73  38 M 0.73  39 D 0.73  40 E 0.73  41 S 0.73  42 H 0.73  43 I 0.73  44 L 0.73  45 E 0.73  46 K 0.73  47 M 0.73</p>
--	---

48 A	0.73
49 A	0.73
50 E	0.73
51 A	0.73
52 G	0.73
53 K	0.73
54 K	0.73
total: 28 residues	
-----	
Final Results (k = 9/23):	
44.4 %: endoplasmic reticulum	
22.2 %: Golgi	
11.1 %: plasma membrane	
11.1 %: vesicles of secretory system	
11.1 %: extracellular, including cell wall	
>> indication for CG139008-01 is end (k=9)	

A search of the NOV3a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 3D.

5

Table 3D. Geneseq Results for NOV3a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM51697	Human new lipid binding protein 2 - Homo sapiens, 454 aa. [WO200179493-A1, 25-OCT-2001]	1..454 1..454	454/454 (100%) 454/454 (100%)	0.0
AAB47337	FCTR14 - Homo sapiens, 454 aa. [WO200146231-A2, 28-JUN-2001]	1..454 1..454	454/454 (100%) 454/454 (100%)	0.0
ABB08898	Human BPIL 325-3 SEQ ID NO 35 - Homo sapiens, 454 aa. [WO200136478-A2, 25-MAY-2001]	1..454 1..454	454/454 (100%) 454/454 (100%)	0.0
ABB08899	Human BPIL 325-4 SEQ ID NO 45 - Homo sapiens, 453 aa. [WO200136478-A2, 25-MAY-2001]	1..444 1..443	442/444 (99%) 443/444 (99%)	0.0
ABG10878	Novel human diagnostic protein #10869 - Homo sapiens, 455 aa. [WO200175067-A2, 11-OCT-2001]	1..454 1..455	441/455 (96%) 444/455 (96%)	0.0

In a BLAST search of public sequence databases, the NOV3a protein was found to have homology to the proteins shown in the BLASTP data in Table 3E.

<b>Table 3E. Public BLASTP Results for NOV3a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV3a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
CAC50178	Sequence 27 from Patent WO0146231 - Homo sapiens (Human), 454 aa.	1..454 1..454	454/454 (100%) 454/454 (100%)	0.0
Q8NFQ5	Bactericidal/permeability-increasing protein-like 3 - Homo sapiens (Human), 453 aa.	1..444 1..443	442/444 (99%) 443/444 (99%)	0.0
Q05704	Potential ligand-binding protein - Rattus rattus (Black rat), 470 aa (fragment).	59..444 73..463	130/395 (32%) 229/395 (57%)	3e-57
CAD12150	Sequence 3 from Patent WO0179269 - Homo sapiens (Human), 637 aa.	59..444 241..630	125/394 (31%) 222/394 (55%)	2e-52
CAC18887	DJ726C3.5 (ortholog of potential ligand_binding protein RY2G5 (Rat)) - Homo sapiens (Human), 469 aa (fragment).	59..444 73..462	125/394 (31%) 222/394 (55%)	2e-52

- 5 Pfam analysis indicates that the NOV3a protein contains the domains shown in the Table 3F.

<b>Table 3F. Domain Analysis of NOV3a</b>			
<b>Pfam Domain</b>	<b>NOV3a Match Region</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
LBP_BPI_CETP_C	291..429	41/140 (29%) 95/140 (68%)	1.3e-11

**Example 4.**

The NOV4 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 4A.

Table 4A. NOV4 Sequence Analysis			
	SEQ ID NO: 19	765 bp	
NOV4a, CG145877-01 DNA Sequence	TCGCCCTTCATGGTGATGTCCCAGGCCACCTACACGTTCTCAGTGCTTCGCCGGCTTCTGGC TCATCTGGGGTCTCATCGTCTGCTGCTGCTTCTGCAGCTTCTGCGCCGCCGCTCAAACG GCGCCAGGAGGAGCGACTGCGCGAGCAGAACCTGCGCGCCCTAGAGCTGGAGCCCCCTCGAACTC GAGGGCAGTCTGGCCGGGAGCCCCCGGGCTGGCGCCGCCGAGCCACCACACCGTAGCC GCCTGGAGGCGCGGCTCAGCGCACTCGCATCCGCACGTGCACGTGCACCCGCCGCTACGCA CCTGTCGGTGCCGCCACGGCCCTGGAGCTACCCGCGCCAAGCGGAATCGGACATGTCAAACCA CCGTGTTACGAAGAGGCGGTGCTGATGGCAGAGCCGCCGCCGCTATAGCGAGGTGCTCACGG ACACGCGCGGCTCTACCGCAAGATCGTCACGCCCTTCTGAGTCGCCGCGACAGCGCGGAGAA GCAGGAGCAGCCGCTCCAGCTACAAGCCGCTCTTCTGGACCGGGGCTACACCTCGGGCGCTG CACCTGCCCAGCGCCCTCGGCCCGCGCCGCTGCCAGCCCTCTGCTGCAGGCGGACCGTG GCCGCCGGTCTTCCCCAGCTGGACCGACTCAGAGCTCAGCAGCGCGAGCCCCTGGAGCACGG AGCTTGGCGTCTGCCGCTCTCCATCCCCTTGTTCGGGAGGACTACAGCCGTATAGAGGGGC		
	ORF Start: ATG at 10		ORF Stop: TAG at 757

5

	SEQ ID NO: 20	249 aa	MW at 28180.1kD
NOV4a, CG145877-01 Protein Sequence	MVMSQATYTFLTCTFAGFWLIWGLIVLLCCFCSFLRRRLKRRQEERLREQNLRLELEPLELEGS LAGSPFGLAPPQPPPHRSRLAAPAHASHPHVHVHPPPTHLSVPPRPWSYPRQAESDMSKPPCY EEAVLMAEPPPPYSEVLTDTRGLYRKIVTFFLSRRDSAEKQEQPPPSYKPLFLDRGYTSALHLP SAPRPAPPCPALCLQADRGRVFPFSWTDSELSSREPLEHGAWRLFPVSIPLFGRTTAV		

Further analysis of the NOV4a protein yielded the following properties shown in

10 Table 4B.

Table 4B. Protein Sequence Properties NOV4a	
SignalP analysis:	Cleavage site between residues 35 and 36
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 0; pos.chg 0; neg.chg 0  H-region: length 34; peak value 11.41  PSG score: 7.01</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): 0.65  possible cleavage site: between 31 and 32</p> <p>&gt;&gt;&gt; Seems to have a cleavable signal peptide (1 to 31)</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 32  Tentative number of TMS(s) for the threshold 0.5:  0</p>

	<p>number of TMS(s) .. fixed  PERIPHERAL Likelihood = 8.33 (at 233)  ALOM score: 8.33 (number of TMSs: 0)</p> <p>MTOP: Prediction of membrane topology (Hartmann et al.)  Center position for calculation: 15  Charge difference: 4.0 C( 5.0) - N( 1.0)  C &gt; N: C-terminal side will be inside</p> <p>&gt;&gt;&gt;Caution: Inconsistent mtop result with signal peptide  MITDISC: discrimination of mitochondrial targeting seq  R content: 5 Hyd Moment(75): 4.61  Hyd Moment(95): 3.22 G content: 2  D/E content: 1 S/T content: 5  Score: -0.42</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  R-2 motif at 51 RRQ EE`</p> <p>NUCDISC: discrimination of nuclear localization signals  pat4: none  pat7: none  bipartite: none  content of basic residues: 11.6%  NLS Score: -0.47</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals: none</p> <p>SKL: peroxisomal targeting signal in the C-terminus: none</p> <p>PTS2: 2nd peroxisomal targeting signal: none</p> <p>VAC: possible vacuolar targeting motif: none</p> <p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif:  type 1: none  type 2: none</p> <p>NMYR: N-myristoylation pattern : none</p> <p>Prenylation motif: none</p> <p>memYQRL: transport motif from cell surface to Golgi: none</p> <p>Tyrosines in the tail: none</p> <p>Dileucine motif in the tail: none</p> <p>checking 63 PROSITE DNA binding motifs: none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p>
--	---

	<p>none</p> <p>NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination Indication: nuclear Reliability: 94.1</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions total: 0 residues</p> <p>-----</p> <p>Final Results (k = 9/23):</p> <p>43.5 %: mitochondrial 43.5 %: nuclear 13.0 %: extracellular, including cell wall</p> <p>&gt;&gt; indication for CG145877-01 is mit (k=23)</p>
--	--

A search of the NOV4a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 4C.

5

Table 4C. Geneseq Results for NOV4a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG08144	Novel human diagnostic protein #8135 - Homo sapiens, 436 aa. [WO200175067-A2, 11-OCT-2001]	12..77 312..376	35/66 (53%) 38/66 (57%)	3e-10
ABG27250	Novel human diagnostic protein #27241 - Homo sapiens, 406 aa. [WO200175067-A2, 11-OCT-2001]	65..202 23..150	43/140 (30%) 54/140 (37%)	4e-06
AAG67355	Amino acid sequence of a rat N-WASP protein - Rattus rattus, 501 aa. [WO200144292-A2, 21-JUN-2001]	66..140 294..373	32/80 (40%) 36/80 (45%)	1e-05
AAM52319	Rat N-WASP protein - Rattus rattus, 501 aa. [WO200171356-A2, 27-SEP-2001]	66..140 294..373	32/80 (40%) 36/80 (45%)	1e-05
AAW46890	Rat Neural-Wiskott-Aldrich syndrome protein - Rattus sp, 501 aa. [JP10072494-A, 17-MAR-1998]	66..140 294..373	32/80 (40%) 36/80 (45%)	1e-05

In a BLAST search of public sequence databases, the NOV4a protein was found to have homology to the proteins shown in the BLASTP data in Table 4D.

Table 4D. Public BLASTP Results for NOV4a				
Protein Accession Number	Protein/Organism/Length	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BTA7	Hypothetical protein - Homo sapiens (Human), 253 aa.	1..249 1..253	245/253 (96%) 246/253 (96%)	e-147
Q8TB68	Hypothetical protein MGC10772 - Homo sapiens (Human), 274 aa.	1..249 1..274	248/274 (90%) 248/274 (90%)	e-146
Q8WU53	Similar to hypothetical protein MGC10772 - Homo sapiens (Human), 274 aa.	1..249 1..274	247/274 (90%) 247/274 (90%)	e-145
P13983	Extensin precursor (Cell wall hydroxyproline-rich glycoprotein) - Nicotiana tabacum (Common tobacco), 620 aa.	69..202 302..414	38/134 (28%) 49/134 (36%)	2e-06
Q94ES6	Nodule extensin - Pisum sativum (Garden pea), 181 aa.	64..203 31..169	40/144 (27%) 53/144 (36%)	7e-06

5

#### Example 5.

The NOV5 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 5A.

Table 5A. NOV5 Sequence Analysis			
	SEQ ID NO: 21	1126 bp	
NOV5a, CG151161-02 DNA Sequence	GGCACGAGGCCCGCGCGCGGGGCGCCAGGCCACTGGGCTCCGCGGAGCCAGCGAGAGGTCTG CGCGGAGTCTGAGCGGCGCTCGTCCCGTCCCAAGGCCGACGCCAGCACGCGGTCATGGCCCCG CAGCGGCGACGGGGGCGAGCACCTGCCAGTGGCTTCTCGGTCTTCAACACCTTGCCGACTT GCTCTTCATCTTTGAGTTTATCTTCGGGGGCGCTGGTGTGGATCCTGGTGGCCCTCTCCCTGGT CCCTGGCCCTGGTCCAGGGCTGGGTGATGTTCTGTCTGTGTTCTGCTTCGTGGCCACCACCA CCTTGATCATCTGTACATAATTGGAGCCACGGTGGAGAGACTTCCTGGGTACCTTGGACGC AGCCTACCACTGCACCGCTGCCCTCTTTACCTCAGCGCCTCAGTCCTGGAGGCCCTGGCCACC ATCAGATGCAAGACGGCTTACCTACAGGCACTACCATGAAAACATTGCTGCCGTGGTGTCT CTTACATAGCCACTCTGCTCTACGTGGTCCATGCGGTGTTCTCTTTAATCAGATGGAAGTCTTC ATAAAGCCGAGTAGAACTTGAGCTGAAAACCCAGATGGTGTAACTGGCCGCCCCACTTTCCG GCATAACTTTTAGAAAAACAGAAATGCCCTTGATGGTGGAAAAAGAAAAACACACCCCCCA CTGCCCAAAAAAAGCCCTGCCCTGTGCTCGTGGGTGCTGTGTTACTCTCCCGTGTGCCT TCGCGTCCGGGTGGGAGCTTGCTGTGTCTAACCTCCAACCTGCTGTGCTGTCTGTAGGGTCAC CTCCTGTTGTGAAAGGGGACCTTCTGTTCGGGGGTGGGAAGTGGCGACCGTGACCTGAGAAG GAAAGAAAGATCCTCTGCTGACCCCTGGAGCAGCTCTCGAGAACTACCTGTTGGTATTGTCCAC		

	AAGCTCTCCCGAGCGCCCCATCTTGTGCCATGTTTAAAGTCTTCATGGATGTTCTGCATGTCAT GGGGACTAAAACTCACCCAAACAGATCTTCCAGAGGTCCATGGTGGAGACGATAACCTGTGA AATACTTTATAAAATGTCTTAATGTTCAAAAAAAAAA		
	ORF Start: ATG at 119		ORF Stop: TAA at 578

	SEQ ID NO: 22	153 aa	MW at 16713.3kD
NOV5a, CG151161-02 Protein Sequence	MAPAAATGGSTLPSGFSVFTLPDLLEFIFEGGLVWILVASSLVWPLVQGWVMFVSFVCFV ATTTLIILYIIIGAHGGETSWVTLDAAVHCTAALFYLSASVLEALATITMQDGFTYRHYHENIAA VVFSLYIATLLYVVHAVFSLIRWKSS		

5

	SEQ ID NO: 23	464 bp	
NOV5b, CG151161-01 DNA Sequence	GGCACGAGGCCCGCGCGCGGGGCGCCAGGCCACTGGGCTCCGCGAGCCAGCGAGAGGCTCTG CGCGGAGTCTGAGCGGCGCTCGTCCCGTCCCAAGGCCGACGCCAGCAGCCGTCATGGCCCCCG CAGCGGCGACGGGGGGCAGCACCTGCCAGTGGCTTCTCGGTCTTCACCACCTTGCCCGACTT GCTCTTCATCTTTGAGTTTGACGCAACCTACCACTGCACCGCTGCCCTCTTTACCTCAGCGCC TCAGTCCTGGAGGCCCTGGCCACCATCACGATGCAAGACGGCTTCACTACAGGCACTACCATG AAAACATTGCTGCCGTGGTGTCTCTACATAGCCACTCTGCTCTACGTGGTCCATCGGCTGTT CTCTTTAATCAGATGGAAGTCTTCAAAAGCCGAGTAGAACTTGAGCTGAAAACCCAGATGGT GTTAACTGGCCGCCCC		
	ORF Start: ATG at 119		ORF Stop: TAA at 410

	SEQ ID NO: 24	97 aa	MW at 10651.1kD
NOV5b, CG151161-01 Protein Sequence	MAPAAATGGSTLPSGFSVFTLPDLLEFIFEDATYHCTAALFYLSASVLEALATITMQDGFTYR HYHENIAAVVFSLYIATLLYVVHAVFSLIRWKSS		

10

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 5B.

Table 5B. Comparison of NOV5a against NOV5b.		
Protein Sequence	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV5b	1..153 1..97	94/153 (61%) 94/153 (61%)

15

Further analysis of the NOV5a protein yielded the following properties shown in Table 5C.

Table 5C. Protein Sequence Properties NOV5a	
SignalP analysis:	Cleavage site between residues 67 and 68

PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 0; pos.chg 0; neg.chg 0  H-region: length 23; peak value 8.79  PSG score: 4.39</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): -1.89  possible cleavage site: between 53 and 54</p> <p>&gt;&gt;&gt; Seems to have a cleavable signal peptide (1 to 53)</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 54  Tentative number of TMS(s) for the threshold 0.5:  3  INTEGRAL Likelihood = -7.64 Transmembrane  55 - 71  INTEGRAL Likelihood = -0.90 Transmembrane  95 - 111  INTEGRAL Likelihood = -4.30 Transmembrane  129 - 145  PERIPHERAL Likelihood = 10.08 (at 74)  ALOM score: -7.64 (number of TMSs: 3)</p> <p>MTOP: Prediction of membrane topology (Hartmann et al.)  Center position for calculation: 26  Charge difference: 0.0 C(-1.0) - N(-1.0)  N &gt;= C: N-terminal side will be inside</p> <p>&gt;&gt;&gt; membrane topology: type 3a</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 0 Hyd Moment(75): 2.25  Hyd Moment(95): 2.24 G content: 3  D/E content: 1 S/T content: 7  Score: -5.30</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  cleavage site motif not found</p> <p>NUCDISC: discrimination of nuclear localization signals  pat4: none  pat7: none  bipartite: none  content of basic residues: 2.0%  NLS Score: -0.47</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals:  KKXX-like motif in the C-terminus: RWKS</p> <p>SKL: peroxisomal targeting signal in the C-terminus:  none</p> <p>PTS2: 2nd peroxisomal targeting signal: none</p> <p>VAC: possible vacuolar targeting motif: none</p>
-----------------------	--

	<p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif:  type 1: none  type 2: none</p> <p>NMYR: N-myristoylation pattern : none</p> <p>Prenylation motif: none</p> <p>memYQRL: transport motif from cell surface to Golgi:  none</p> <p>Tyrosines in the tail: none</p> <p>Dileucine motif in the tail: none</p> <p>checking 63 PROSITE DNA binding motifs: none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p> <p>checking 33 PROSITE prokaryotic DNA binding motifs:  none</p> <p>NNCN: Reinhardt's method for Cytoplasmic/Nuclear  discrimination  Indication: cytoplasmic  Reliability: 94.1</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions  total: 0 residues</p> <p>-----</p> <p>Final Results (k = 9/23):</p> <p>66.7 %: endoplasmic reticulum  33.3 %: mitochondrial</p> <p>&gt;&gt; indication for CG151161-02 is end (k=9)</p>
--	--

A search of the NOV5a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded

5 several homologous proteins shown in Table 5D.

Table 5D. Geneseq Results for NOV5a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV5a	Identities/	Expect Value

		Match Residues	the Matched Region	
ABB50292	T cell differentiation protein Mal ovarian tumour marker protein, #74 - Homo sapiens, 153 aa. [WO200175177-A2, 11-OCT-2001]	1..153 1..153	153/153 (100%) 153/153 (100%)	5e-85
AAP80929	Sequence of human T-cell protein designated MAL - Homo sapiens, 153 aa. [WO8807549-A, 06-OCT-1988]	1..153 1..153	150/153 (98%) 151/153 (98%)	3e-82
AAP81879	Sequence of full-length human T-cell protein derived from mature T cells - Homo sapiens, 153 aa. [WO8807549-A, 06-OCT-1988]	1..153 1..153	150/153 (98%) 151/153 (98%)	3e-82
AAU85517	Clone #18966 of lung tumour protein - Homo sapiens, 148 aa. [WO200204514-A2, 17-JAN-2002]	3..143 2..142	60/141 (42%) 91/141 (63%)	8e-28
AAB76862	Human lung tumour protein related protein sequence SEQ ID NO:338 - Homo sapiens, 148 aa. [WO200100828-A2, 04-JAN-2001]	3..143 2..142	60/141 (42%) 91/141 (63%)	8e-28

In a BLAST search of public sequence databases, the NOV5a protein was found to have homology to the proteins shown in the BLASTP data in Table 5E.

Table 5E. Public BLASTP Results for NOV5a				
Protein Accession Number	Protein/Organism/Length	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P21145	Myelin and lymphocyte protein (T-lymphocyte maturation-associated protein) - Homo sapiens (Human), 153 aa.	1..153 1..153	153/153 (100%) 153/153 (100%)	1e-84
Q64349	Myelin and lymphocyte protein (T-lymphocyte maturation-associated protein) (17 kDa myelin vesicular protein) (MVP17) (NS 3) - Rattus norvegicus (Rat), 153 aa.	1..153 1..153	136/153 (88%) 147/153 (95%)	2e-77
O09198	Myelin and lymphocyte protein (T-lymphocyte maturation-associated protein) - Mus musculus (Mouse), 153 aa.	1..153 1..153	133/153 (86%) 145/153 (93%)	2e-75

Q28296	Myelin and lymphocyte protein (T-lymphocyte maturation-associated protein) (VIP17 proteolipid) - <i>Canis familiaris</i> (Dog), 153 aa.	1..153 1..153	135/153 (88%) 146/153 (95%)	2e-75
Q9D2R2	Myelin and lymphocyte protein; T-cell differentiation protein - <i>Mus musculus</i> (Mouse), 97 aa.	1..153 1..97	84/153 (54%) 91/153 (58%)	4e-34

**Example 6.**

The NOV6 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 6A.

Table 6A. NOV6 Sequence Analysis			
	SEQ ID NO: 25	4801 bp	
NOV6a, CG155653-01 DNA Sequence	CCGCTGCGGGCTCGGGCGCCGACGCGCGCGCCGAGCCCTGGACGAGGCCACGGAGCCGC TCGCCCCGAGCCAGCCGCGCCGATGTCTCAAAATGGAGGCAGCGGGGCGGGCGGCGTGAAGAAA GCGGCGCTGTGGGCGCGGGAGTAGGGGCCCGGGCGGAGGCGGTGGCGGGATGGGGCTGCTGCTC ATGATCCTGGCGTCGGCCGTGCTGGGTTCTTCTCACGCTCCTCGCCAGTTCTTCTGCTGT ACCGCAGACAGCCCGAGCCGCGCGGACGAGGCGCGCCGCGGGCGAGGGCTTCCGCTACAT CAAGCCAGTGCCGGGCTGCTCCTAAGGGAGTACCTTTATGGCGGCGCGGGATGAGGAGCCC TCCGGAGCGGCCCTGAGGGCGCGCAGCCCCACCGCGGCCCGGAGACCCCGCCCGCCGA CGCGGGAGACTTGCTACTTCTCAACGCCACCATCTTATCTGTTCCGGGAGTTGCGGGACAC CGCGCTGACCGCGCTGGGTCAACAAGAGATCAAGGTGAGTTCCGAGGAGCTGCTGCAGACC AAGACGGCCGGCGCTGCTGGAGGGGTGAGCCTGCGGACGTGTTCTTGGCGAGACGGTGC CCTTCATCAAGACATCCGGCTCGTGCGGCCAGTCTGCGCCCTCGGCCACCGGGGAGCCGATGG CCCTGAAGGGGAGGCGCTGCGCGCGCCTGCCCGAGGAGCTGGCCCTTCGAGGCGGAGTGGAG TACAACGGGGGCTTCCACCTGGCCATCGACGTGGACCTGGTCTTCGGCAAGTCCGCTACTTGT TTGTCAAGCTGTCCCGCTGGTGGGAAGGCTGCGCTTGGTCTTTACGCGCGTCCCTTCAACCA CTGGTTCTTCTCCTCGTGAAGACCCGCTGATCGACTTCGAGGTGCGCTCCAGTTTGAAGGG CGGCCCATGCCCGAGCTCACCTCCATCATCGTCAACCAGCTCAAGAAGATCATCAAGCGCAAGC ACACCCTACCGAATTACAAGATCAGGTTTAAGCCGTTTTTTCCATACCAGACCTTGAAGGATT TGAAGAAGATGAAGAGCATATCCATATACAACAATGGGCACCTACTGAAGGCCGTCTTAAAGTT ACGTTGTTAGAATGTAGCAGTTACTCATTTTGGATCCTATGACAGAGAGGCAAATGTTTATT GCACACTTGAGTTAAGCAGTAGTGTGTTGGGAAGAAAAACAGAGGAGTTCTATTAAAGCGTTGA ATTAATAAAAGGAAATTACAAAGTGTGGACTTACACTTCGTCTTGTCCAGTCAACTGATGGG TATGCTGGGCAGCTCATATTGAAACTGTGGCTCCAAACTCGCCTGTGCAATTGCAGATCTTC AGCGGGGAGATCGACTTATCGCCATTGGAGGTGTGAAATCAATCAACTGCAAGTGTGAA GCTTATCAAGCAGGCTGGTACCAGTCTGCTGTACTATGAAAGGCTGTGGCCAGAGTAAT CAAGGTGCACTGCTGCAAGATAACTTTGGCCAGTTGGAAGAAAACCTTTTGTCAAGCTCATGCC AATCGGGTTATGAAGAGGAAGCTGCCGGTTGACAGTAGATACTGAAGTAGAGAGCTGGATTCT TGAATTTGAAGACTTGGCAAGTGATGTGAGAGCACAATAAGATTCAAGATGAGGCACAACTCA TTAAGTCATAGTCCCAAACGTGTTCCAACAACACTTTCTATTAAACCCCTTGGAGCTATATCAC CAGTTTAAACCGTAAATAGCTGTAGGAAGTCAACCACTACCAACGAAAATTAGTCCAAAGA TGGAAATAAACCTCCACCCCTAAAACTTCTGAGATAACAGACCCAGCACAAGTGTCAAAACCA ACCAAGGATCTGCTTTCAAACCTGTGCCACCAAGCAAGCAAGGAAAAGTTCTTATGCTT CCGCCGATGCTCCAAATCAGGCAGAACAGATGTTCTGTTGAAAAGCCAGAGAAGGTGGTGCC ACCTCCTCTTGTAGATAAATCTGTGAAAAGCAAGCAAAAATGTGGATGCCATAGACGATGCA GCTGCACCTAAGCAATTTTAGCAAAGCAAGAGTGGCCAAAGATGTCACTTCAGAAACTTCT GCCCTACTAAGGACAGTTGCGACGACCGTCAACATGCGGAATCATCAAGAAATTTTATGCTAA TAAGTAGGAAAATGGACAAGAACCAGAGCATCCTGTTGTTGACATAGAAGCCTGTACAGG TACTTAAACATTGCATTGTTGGTGAGGGATCCTTTCAAGTTGGGAGGTCTCATGTTTGGGGC ATGTTAGTTTAAACCTGAAGATGTGGCTTAGGATGCTAGCTACTCAACACGGAATACCT TTCCAAATTGAGACTGGAAGCCCCCTCACCTAAGGCTATAGTCACTAGAACCGCACTACGCAAT		

	<p>CTGAGTATGCAAAAGGGATTCAATGACAAATTTGCTATGGTGACATTACTATTCACTTCAAAT  ATTTGAAAGGAAGGAGAATCAGACCCATGTAGTTACTAACGTAGAAAAAGAAAGACCCCA  TTTGGTTGAAGAAGTTTCTGTTCTCCCTAAAGAGGAGCAATTTGTTGGACAGATGGGTTTAAACA  GAAAAACAACACAGTTTTTTCAGGATACCTAGTTCCAGAACCCACATGGTGTGACTACTGTAAAGA  AAAAAGTTTGGACTAAAGCAGCTTCCAGTGTATGTTTTGTGCTTATGTTTGGCCATAAAAAATG  TCAAGAAAAGTGTCTAGCTGAGACTTCTGTTTGTGGAGCAACTGATAGGCGAATAGACAGGACA  CTGAAAAACCTTAGGCTGGAAGGACAGGAAACCTCTTAGGCCTGCCTCCTCGTGTGATGCTG  AAGCTAGCAAGTCAGTCAATAAAACACAGGTTTGACAAGGCATATATCAATACTAGTTCTCG  TTTATTAAATTTGCGTCAAGTCTCTAAAACTCGCCTTTCTGAACCAGGAACCGATCTCGTAGAA  CCTTCACCAAAACACACACCCCAACCGTCAGACAACGAAGGCACTGACACGGAGGTCTGTGGTC  CAAAACAGTCTTTCTAAACGGGGAAACAGCACAGGAATAAAGTTAGTGAGAAAAGAGGGTGGTCT  GGATGACAGTGTTTTCATTGCAAGTTAAAGAAATGGTTCGTGATCTGTACAGGGGCTTGCCATACA  GAGGAAAGGATCCAGAACTAGAGTTTCATGTTGGATAAGCTACAGAAATGAAATGATCAGGAGT  TGGAAACACAATAATTCCTTGTGTAGAGAAAGAAAGAGACAACTGATACAAGGAAAAATCACT  TCTTCTGCTGCCTTAGCTAAATCAGGTGAAAGGCTACAAGCTCTAACACTTCTTATGATTACAC  TACAGAGCAGGCATTGAAGATATAGAACTTTAGAAAGTCTGCTTTAGACCAGCACTCCAAAA  AAATAAGCAAGTACACAGATGATACAGAAGAAGACCTTGATAATGAAATAAGCCAACATAAGA  CTCTCAGCCATTACAGCAGCATATCAGATGACTTATTTGGCCCATCCGAGTCTGTGTAGCAGACA  GGTCTATTTAAACTTTCAAATGAAACAGGGTAAAGTTGCATCTAAAGTACCACAGATACAACCAT  GTTAAATCCTCGTATGCACTCTGGCCTGCTTCTCCAGTTACTTGTGTGTAAGAAACAAAAAT  GAGAAAGGTTGTTTTCCAGTAAAAACATGACCAGCTTACTAATTGGTTGTTTTGGATTGCAATTT  ATAGCTATGCTTTTTTGGGTTTATACCTGGGAATTTATTTTTACTAAATTTTAACTTTTCTAA  TTATGTAATATGTAAGCTAGCTTTTCATGTTTATGTATGTATGGTGTCCCTTGTGTTATTTT  TCTTCCTCTTGGTTTTTGAATTAGTGTAAATAGAATACTGTCTGGATTCTTAAATATTTTCA  TTTTCCATCATGGTTTATAACAAATTTGCTGCTATGCCCAAACCTGACAACTGACAACTCAGGAGGAA  CAGGTTTTGAATCTTTCTTTTGTGTATGAAGTTTATCGTCTCTACTTGTCTTGAGATTTTTGTT  ATTTTGGGGGTTTTGGGGGTGCTTTTTGTTTTGTTTTGCCAAATGTAACATGAAGCAGATGCT  GCAGCTTTAGTCTGTTATGCTGATTTAGTAAAAAAATTTTACATATATTTGCTTGTCTTTCG  ATGCTTCTGTGAAATTTTTTCTAAAGCTTTTGTGCTGCTATGGTAAAAATATGGTGATTAA  TTTGAAGAGCTTACATTGAAGACAATGTAATAGGAAATAAATGTAGATGCACTTGGTCAAGA  ATTTGTAGAGAGGATAACAAGACTTAATTAAGTAAAAACAGTAACATAGCATTTTGAAATATG  ATCTTTTAAATATTTGATGCTTTCTTTTAAATGGAATTTAAATTTATAATTTAAAGTTTAA  ACATTTATGATAATTTTCTCATCAGTTTCTCCATAGGAAATAAAGCATGTGAAGGGTATTTA  AAGTTTGGAGGACTCTTTTAAATAGCTGTGTGTGATACTAGTTTGGGCTGGTTTTGTTTTA  GAAAAACATTTTCATGTAGGAGTATCTGTGAAGGAAAGGAATCATGCAAAATATATCTTTTG  CTTTGGCGTCTTACAGTTGTAAGGAATGGTGATCATTTCTGAATACCTCTGTAGTGAGTATTCA  T</p>
	<p>ORF Start: ATG at 178</p>
	<p>ORF Stop: TAG at 3640</p>

SEQ ID NO: 26	1154 aa	MW at 128561.7kD
<p>NOV6a, CG155653-01 Protein Sequence</p>	<p>MGLLLMILASAVLGSFLTLAQFFLLYRRQPEPPADEAARAGEGFRYIKPVPGLLLREYLYGGG  RDEEPSGAAPPEGATPTAAPETPAPPTRETCTYFLNATILFLFRELRTALTRRWVTKKIKVEFE  ELLQTKTAGRLLEGLSLRDVFLGETVFPFIKIRLVRPVVPSATGEPDGPGEALPAACPEELAF  EAEVEYNGGFHLAIDVDLVFGKSAFLFVKLSRVVGRRLRVFTRVPFTHWFFSFVEDPLIDFEVR  SQFEGRPMPLTSIIIVNQLKKIIKRKHTLPNYKIRFKPFFPYQTLOGFEEDEEHIHQWALTE  GRLKVTLLCESTRLLIFGSYDREANVHCTLELSSSVWEEKQORSSIKTVELIKGNLQSVGLTLRLV  QSTDGYAGHVIETVAPNSPAIADLQRGDRLIAIGGVKITSTLQVLKLIKQAGDRVLVYYER  VGQSNQGAVLQDNFGQLEENFLSSSCQSGYEEEAAGLTVDTESRELDSEFEDLASDVRAQNEFK  DEAQSLSHSPKRVPTTSLIKPLGAI SPVLNRKLAVGSHPLPKIQSKDGNKPPPLKTSEITDPA  QVSKPTQGSFAKPPVPPRPQAKVPLPSADAPNQAEPDVLVEKPEKVPPPLVDKSAEKQAKNVD  AIDDAAPKQFLAKQEVAKDVTSETSCPTKDDSDRQTWESSEILYRNKLKQWTRTRASCFLDI  EACHRYLNIALWCRDPFLKGLGLCLGHVSLKLEDVALGCLATSNTYLSKRLRLAPSPKAIIVTR  TALRNLSMQKGFNDKFCYGDITTHFKYLKEGESDHHVTVNVEKEKEPHLVEEVSVLPKEEQFVG  QMGLTENKHSFQDTQFQNPWCDCYCKKVVTKAASQCMFCAYVCHKCKQEKLAETSVCGATDR  RIDRTLKNLRLLEGQETLLGLPPRVDAEASKSVNKTGLTRHIINTSSRLNLRQVSKTRLSEPG  TDLVEPSPKHTPNTSDNEGSDTEVCGPNPSKRGNSTGIKLVRKEGGLDDSVFIIVKEIGRDLY  RGLPTEERIQLKLEFMDKLQNEIDQLEHNNSLVREEKETTDTRKSLLSAALAKSGERLQALT  LLMIHYRAGIEDIETLESLSLDQHSKKISKYTDDTEEDLDNEISQLIDSQPFSSISDDLFGPSE  SV</p>	

Further analysis of the NOV6a protein yielded the following properties shown in Table 6B.

Table 6B. Protein Sequence Properties NOV6a	
SignalP analysis:	Cleavage site between residues 22 and 23
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 0; pos.chg 0; neg.chg 0  H-region: length 27; peak value 11.55  PSG score: 7.15</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): 0.74  possible cleavage site: between 14 and 15</p> <p>&gt;&gt;&gt; Seems to have a cleavable signal peptide (1 to 14)</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 15  Tentative number of TMS(s) for the threshold 0.5:  0  number of TMS(s) .. fixed  PERIPHERAL Likelihood = 1.38 (at 204)  ALOM score: 1.38 (number of TMSs: 0)</p> <p>MTOP: Prediction of membrane topology (Hartmann et al.)  Center position for calculation: 7  Charge difference: -1.0 C( 0.0) - N( 1.0)  N &gt;= C: N-terminal side will be inside</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 2 Hyd Moment(75): 1.54  Hyd Moment(95): 1.27 G content: 2  D/E content: 1 S/T content: 3  Score: -4.42</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  R-2 motif at 39 RRQ PE</p> <p>NUCDISC: discrimination of nuclear localization signals  pat4: KRKH (3) at 280  pat7: none  bipartite: none  content of basic residues: 12.5%  NLS Score: -0.29</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals: none</p> <p>SKL: peroxisomal targeting signal in the C-terminus:  none</p> <p>PTS2: 2nd peroxisomal targeting signal: none</p> <p>VAC: possible vacuolar targeting motif: found</p>

	<p>TLEN at 284</p> <p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif:  type 1: none  type 2: none</p> <p>NMYR: N-myristoylation pattern : none</p> <p>Prenylation motif: none</p> <p>memYQRL: transport motif from cell surface to Golgi:  none</p> <p>Tyrosines in the tail: none</p> <p>Dileucine motif in the tail: none</p> <p>checking 63 PROSITE DNA binding motifs: none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p> <p>checking 33 PROSITE prokaryotic DNA binding motifs:  none</p> <p>NNCN: Reinhardt's method for Cytoplasmic/Nuclear  discrimination  Indication: nuclear  Reliability: 55.5</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions</p> <p>1028 T 1.00  1029 E 1.00  1030 E 1.00  1031 R 1.00  1032 I 1.00  1033 Q 1.00  1034 K 1.00  1035 L 1.00  1036 E 1.00  1037 F 1.00  1038 M 1.00  1039 L 1.00  1040 D 1.00  1041 K 1.00  1042 L 1.00  1043 Q 1.00  1044 N 1.00  1045 E 1.00  1046 I 1.00  1047 D 1.00  1048 Q 1.00  1049 E 1.00  1050 L 1.00  1051 E 1.00  1052 H 1.00  1053 N 1.00  1054 N 1.00  1055 S 1.00</p>
--	--

1056	L	1.00
1057	V	0.99
1058	R	0.99
1059	E	0.99
1060	E	0.94
1061	K	0.94
1062	E	0.94
1063	T	0.83
1064	T	0.83
1065	D	0.83
1066	T	0.83
1067	R	0.67
1068	K	0.67
1069	K	0.67
1070	S	0.67
1071	L	0.67
1072	L	0.67
1073	S	0.67
1074	A	0.67
1075	A	0.67
total: 48 residues		
-----		
Final Results (k = 9/23):		
33.3 %: extracellular, including cell wall		
33.3 %: nuclear		
22.2 %: mitochondrial		
11.1 %: cytoplasmic		
>> indication for CG155653-01 is exc (k=9)		

A search of the NOV6a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 6C.

5

Table 6C. Geneseq Results for NOV6a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM78475	Human protein SEQ ID NO 1137 - Homo sapiens, 1204 aa. [WO200157190-A2, 09-AUG-2001]	1..1154 1..1204	1154/1204 (95%) 1154/1204 (95%)	0.0
AAU99614	Human glioma antigen KU-GB-5 - Homo sapiens, 891 aa. [WO200255695-A1, 18-JUL-2002]	264..1154 1..891	890/891 (99%) 890/891 (99%)	0.0
ABG39902	Human peptide encoded by	436..1147 1..712	711/712 (99%) 712/712 (99%)	0.0

	SEQ ID 29567 - Homo sapiens, 712 aa. [WO200186003-A2, 15-NOV-2001]			
AAM18088	Peptide #4522 encoded by probe for measuring cervical gene expression - Homo sapiens, 712 aa. [WO200157278-A2, 09-AUG-2001]	436..1147 1..712	711/712 (99%) 712/712 (99%)	0.0
AAM70260	Human bone marrow expressed probe encoded protein SEQ ID NO: 30566 - Homo sapiens, 712 aa. [WO200157276-A2, 09-AUG-2001]	436..1147 1..712	711/712 (99%) 712/712 (99%)	0.0

In a BLAST search of public sequence databases, the NOV6a protein was found to have homology to the proteins shown in the BLASTP data in Table 6D.

Table 6D. Public BLASTP Results for NOV6a				
Protein Accession Number	Protein/Organism/Length	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q8NEN9	Similar to PDZ domain proteins - Homo sapiens (Human), 1154 aa.	1..1154 1..1154	1154/1154 (100%) 1154/1154 (100%)	0.0
Q9UFF1	Hypothetical protein - Homo sapiens (Human), 513 aa (fragment).	642..1154 1..513	512/513 (99%) 512/513 (99%)	0.0
Q9VYR9	CG10362 protein (LD34222p) - Drosophila melanogaster (Fruit fly), 1037 aa.	3..494 8..473	148/506 (29%) 242/506 (47%)	2e-46
T20180	hypothetical protein C53B4.4a - Caenorhabditis elegans, 1584 aa.	93..447 203..585	112/387 (28%) 194/387 (49%)	4e-42
Q9U3L2	C53B4.4c protein - Caenorhabditis elegans, 1449 aa.	93..447 68..450	112/387 (28%) 194/387 (49%)	4e-42

5

PFam analysis indicates that the NOV6a protein contains the domains shown in the Table 6E.

Table 6E. Domain Analysis of NOV6a			
Pfam Domain	NOV6a Match Region	Identities/ Similarities for the Matched Region	Expect Value

PDZ	366..448	19/88 (22%) 62/88 (70%)	7.7e-10
DAG_PE-bind	841..888	18/51 (35%) 36/51 (71%)	2.6e-09

**Example 7.**

The NOV7 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 7A.

5

Table 7A. NOV7 Sequence Analysis			
	SEQ ID NO: 27	1157 bp	
NOV7a, CG160093-01 DNA Sequence	CTTCCGCTGTCGGTTTTACCATGGAGCAGCTGAGCTCAGCAAACACCCGCTTCGCCTTGGAC CTGTTCTCTGGCGTTGAGTGAGAACAAATCCGGCTGGAAACATCTTCATCTCTCCCTTCAGCATT CATCTGCTATGGCCATGGTTTTCTGGGGACCAGAGGTAACACGGCAGCACAGCTGTCCAAGAC TTTCATTTCAACACGGTTGAAGAGGTTCAATCAAGATTCCAGAGTCTGAATGCTGATATCAAC AAACGTGGAGCGTCTTATATTCTGAACTTGCTAATAGATTATATGGAGAGAAAACTTACAATT TCCTTCCTGAGTTCCTGGTTTCGACTCAGAAAAACATATGGTGTGACCTGGCCAGTGTGGATT TCAGCATGCCTCTGAAGATGCAAGGAAGACCATAAACAGTGGGTTGATAACATGACCAAACTT GTGCTAGTAAATGCCATCTATTTCAAGGGAAACTGGAAGGATAAATTCATGAAAGAAGCCACGA CGAATGCACCATTCAGATTGAATAAGAAAGACAGAAAACTGTGAAATGATGTATCAGAAGAA AAAATTTGCATATGGCTACATCGAGGACCTTAAGTGCCGTGTGCTGGAACTGCCTTACCAAGGC GAGGAGCTCAGCATGGTCATCTGCTGCCGGATGACATTGAGGACGAGTCCACGGGCTGAAGA AGATTGAGGAACAGTTGACTTTGAAAAAGTTGCATGAGTGGACTAAACCTGAGAATCTCGATT CATTGAAGTTAATGTGAGCTTGCCAGGTTCAAACCTGGAAGAGAGTTACACTCTCAACTCCGAC CTCGCCCGCTAGGTGTGAGGATCTCTTAACAGTAGCAAGGCTGATCTGTCTGGCATGTGAG GAGCCAGAGATATTTTATATCAAAAATGTGCCAAGTCATTTGTGGAAGTGAATGAAGAGGG AACAGAGGCGGCAGCTGCCACAGCAGGCATCGCAACTTTCTGCATGTTGATGCCGAAGAAAAAT TTCACTGCCGACCATCCATTCTTTCTTATTTCGGCATAATTCCTCAGGTAGCATCTTATTCT TGGGAGATTTTCTTCCCTTAGAAGAAAGAGACTGTAGCAATACAAAATCAAGCTTAGTGCT AAGGG		
	ORF Start: ATG at 23		ORF Stop: TAG at 1109

	SEQ ID NO: 28	362 aa	MW at 41001.3kD
NOV7a, CG160093-01 Protein Sequence	MEQLSSANTRFALDLFLALSENNPAIGNIFISPFSSISSAMAMVFLGTRGNTAAQLSKTFHFNTVE EVHSRFQSLNADINKRGASYILKLANRLYGKTYNFLPEFLVSTQRTYGADLASVDFQHASEDA RKTINQWVDNMTKLVLVNAIYFKGNWQDKFMKEATTNAPFRLNKKDRKTVKMMYQKKKFAYGYI EDLKCRVLELPYQGEELSMVILLPDDIEDESTGLKKIEQLTLEKLHEWTKPENLDFIEVNVSL PRFKLEESYTLNSDLARLGVQDLFNSSKADLSGMSGARDIFISKIVHKSFVEVNEEGTEAAAAAT AGIATFCMLMPEENFTADHPFLFFIRHNSGSILFLGRFSSP		

10

	SEQ ID NO: 29	1550 bp	
NOV7b, CG160093-02 DNA Sequence	CGGCGGCCTGTGCGGAGCTGTTTGTGACGGTTTCCAGGCAGCCAGGGCCAGGCCGCGGCTCCTA TCTGCAGCTGCAGGGAGAGAGAGGAGGAACCCCGTGCATTCTAGAGACGATTTCACAACAAGG AGAAATCAGCTTTTGTGCTTACATGCCGAGCAGCCAGCACGGTTCTTCTTTGCTGTCTCTCGGGG GAAATCAGGGCTCTGAGAGTGGAGATCGAGATGGGCTAGTGGGTGGCGGATGGGACGCTGCACG GCCAGACCTGGACTGTGTTTTACCATGGAGCAGCTGAGCTCAGCAAAACCCCGCTTCGCCTT GGACCTGTTCTGGCGTTGAGTGAGAACATCCGGCTGGAAACATCTTCATCTCTCCCTTCAGC ATTTTCATCTGCTATGGCCATGGTTTTCTGGGGACCAAGGTAACACGGCAGCACAGCTGTCCA AGACTTTCCATTTCAACACGGTTGAAGAGGTTCAATCAAGATTCAGAGTCTGAATGCTGATAT CAACAAACGTGGAGCGTCTTATATTCTGAAACTTGCTAATAGATTATATGGAGAGAAAACTTAC AATTTCTCTCCTGAGTTCTTGGTTTCGACTCAGAAAAATATGGTGCTGACCTGGCCAGTGTGG ATTTTCAGCATGCCTCTGAAGATGCAAGGAAGACCATAAACAGTGGGTCAAAGGACAGACAGA AGGAAAAATCCCGAACTGTTGGCTTCGGGCATGGTTGATAACATGACCAAACTTGTGCTAGTA AATGCCATCTATTTCAAGGGAAGTGGAAAGATAAATTCATGAAAGAAGCCACGACGAATGCAC CATTCAGATTGAATAAGAAAGACAGAAAACTGTGAAAATGATGTATCAGAAAGAAAAATTTGC ATATGGCTACATCGAGGACCTTAAGTGCCGTGTGCTGGAAGTGCCTTACCAAGGCGAGGAGCTC AGCATGGTCATCCTGTGCGGATGACATTGAGGACGAGTCCACGGGCTGAAGAAGATTGAGG AACAGTTGACTTTGGAAGTTCATGAGTGGACTAAACCTGAGAATCTCGATTTTCATTGAAGT TAATGTCAGCTTGCCAGGTTCAAACTGGAAGAGAGTTACACTCTCAACTCCGACCTCGCCCGC CTAGGTGTGCAGGATCTCTTTAACAGTAGCAAGGCTGATCTGTCTGGCATGTGAGGAGCCAGAG ATATTTTTATATCAAAAATTGTCCACAAGTCATTTGTGGAAGTGAATGAAGAGGGAAACAGAGGC GGCAGCTGCCACAGCAGGCATCGCAACTTTCTGCATGTTGATGCCCGAAGAAAAATTCAGTGCC GACCATCCATTCTCTTTCTTTATTCGGCATAATTCTCAGGTAGCATCTATTCTTGGGGAGAT TTTCTTCCCTTAGAAGAAAGAGACTGTAGCAATACAAAATCAAGCTTAGTGCTTTATTACCT GAGTTTTTAATAGAGCCAATATGCTTATATCTTTACCAATAAAACCACTGTCCAGAAACAAA AAAAAAAAAAAAAA		
	ORF Start: ATG at 283		ORF Stop: TAG at 1420

	SEQ ID NO: 30	379 aa	MW at 42741.3kD
NOV7b, CG160093-02 Protein Sequence	MEQLSSANTRFALDLFLALSENPNAGNIFISPFSSISSAMAMVFLGTRGNTAAQLSKTFHFNTVE EVHSRFQSLNADINKRGASYILKLANRLYGEKTYNLFPEFLVSTQKTYGADLASVDFQHASEDA RKTINQWVKQTEGKIPELLASGMVDNMTKLVLVNAIYFKGNWKDKFMKEATTNAPFRLNKKDR KTVKMYQKKFAYGYIEDLKCRVLELPYQGEELSMVILLPDDIEDESTGLKKIEQLTLEKLIH EWTKPENLDFIEVNVSLPRFKLEESYTLNSDLARLGVQDLFNSSKADLSGMSGARDIFISKIVH KSFVEVNEEGTEAAAATAGIATFCMLPEENFTADHPFLFFIRHNSSGSILFLGRFPSSP		

5

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 7B.

Table 7B. Comparison of NOV7a against NOV7b.		
Protein Sequence	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV7b	1..362 1..379	362/379 (95%) 362/379 (95%)

10

Further analysis of the NOV7a protein yielded the following properties shown in Table 7C.

Table 7C. Protein Sequence Properties NOV7a	
SignalP analysis:	No Known Signal Sequence Indicated
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 10; pos.chg 1; neg.chg 1  H-region: length 3; peak value 5.12  PSG score: 0.72</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): -4.07  possible cleavage site: between 48 and 49</p> <p>&gt;&gt;&gt; Seems to have no N-terminal signal peptide</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 1  Tentative number of TMS(s) for the threshold 0.5: 1</p> <p>Number of TMS(s) for threshold 0.5: 1  INTEGRAL Likelihood = -2.97 Transmembrane  28 - 44  PERIPHERAL Likelihood = 2.49 (at 314)  ALOM score: -2.97 (number of TMSs: 1)</p> <p>MTOP: Prediction of membrane topology (Hartmann et al.)  Center position for calculation: 35  Charge difference: 3.5 C( 2.5) - N(-1.0)  C &gt; N: C-terminal side will be inside</p> <p>&gt;&gt;&gt;Caution: Inconsistent mtop result with signal peptide  &gt;&gt;&gt; membrane topology: type 1b (cytoplasmic tail 28 to 362)</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 1 Hyd Moment (75): 7.63  Hyd Moment (95): 4.21 G content: 0  D/E content: 2 S/T content: 3  Score: -5.24</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  cleavage site motif not found</p> <p>NUCDISC: discrimination of nuclear localization signals  pat4: none  pat7: none  bipartite: KKDRKTVKMMYQKKKFA at 172  content of basic residues: 11.3%  NLS Score: 0.02</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals: none</p> <p>SKL: peroxisomal targeting signal in the C-terminus: none</p> <p>PTS2: 2nd peroxisomal targeting signal: none</p>

	<p>VAC: possible vacuolar targeting motif: none</p> <p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif:  type 1: none  type 2: none</p> <p>NMYR: N-myristoylation pattern : none</p> <p>Prenylation motif: none .</p> <p>memYQRL: transport motif from cell surface to Golgi:  none</p> <p>Tyrosines in the tail: too long tail</p> <p>Dileucine motif in the tail: found  LL at 214</p> <p>checking 63 PROSITE DNA binding motifs: none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p> <p>checking 33 PROSITE prokaryotic DNA binding motifs:  none</p> <p>NNCN: Reinhardt's method for Cytoplasmic/Nuclear  discrimination  Indication: cytoplasmic  Reliability: 89</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions  total: 0 residues</p> <p>-----</p> <p>Final Results (k = 9/23):</p> <p>34.8 %: nuclear  21.7 %: mitochondrial  21.7 %: cytoplasmic  8.7 %: vesicles of secretory system  4.3 %: vacuolar  4.3 %: peroxisomal  4.3 %: endoplasmic reticulum</p> <p>&gt;&gt; indication for CG160093-01 is nuc (k=23)</p>
--	---

A search of the NOV7a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded

5 several homologous proteins shown in Table 7D.

<b>Table 7D. Geneseq Results for NOV7a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV7a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAB43755	Human cancer associated protein sequence SEQ ID NO:1200 - Homo sapiens, 437 aa. [WO200055350-A1, 21-SEP-2000]	1..362 59..437	362/379 (95%) 362/379 (95%)	0.0
AAR94367	Human elastase inhibitor - Homo sapiens, 379 aa. [WO9610418-A1, 11-APR-1996]	1..362 1..379	362/379 (95%) 362/379 (95%)	0.0
AAR64159	Human elastase inhibitor - Homo sapiens, 379 aa. [US5370991-A, 06-DEC-1994]	1..362 1..379	362/379 (95%) 362/379 (95%)	0.0
AAY55841	Human cytoplasmic antiproteinase-3 protein (CAP-3) - Homo sapiens, 376 aa. [WO9957273-A2, 11-NOV-1999]	1..362 1..376	186/380 (48%) 250/380 (64%)	5e-98
AAR99254	Cytoplasmic antiproteinase-3 protein - Homo sapiens, 376 aa. [WO9624650-A2, 15-AUG-1996]	1..362 1..376	186/380 (48%) 250/380 (64%)	5e-98

In a BLAST search of public sequence databases, the NOV7a protein was found to have homology to the proteins shown in the BLASTP data in Table 7E.

5

<b>Table 7E. Public BLASTP Results for NOV7a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV7a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
P30740	Leukocyte elastase inhibitor (LEI) (Monocyte/neutrophil elastase inhibitor) (M/NEI) (EI) - Homo sapiens (Human), 379 aa.	1..362 1..379	362/379 (95%) 362/379 (95%)	0.0
P05619	Leukocyte elastase inhibitor (LEI) - Equus caballus (Horse), 379 aa.	1..362 1..379	297/379 (78%) 326/379 (85%)	e-169
Q9D154	1190005M04Rik protein (RIKEN cDNA 1190005M04 gene) (Serine protease inhibitor EIA) - Mus musculus (Mouse), 379 aa.	1..362 1..379	291/379 (76%) 330/379 (86%)	e-167

P80229	Leukocyte elastase inhibitor (LEI) (Leucocyte neutral proteinase inhibitor) (LNPI) - Sus scrofa (Pig), 378 aa.	1..362 1..378	291/379 (76%) 332/379 (86%)	e-166
S38962	serpin - pig, 378 aa.	1..362 1..378	291/379 (76%) 330/379 (86%)	e-165

PFam analysis indicates that the NOV7a protein contains the domains shown in the Table 7F.

Table 7F. Domain Analysis of NOV7a			
Pfam Domain	NOV7a Match Region	Identities/ Similarities for the Matched Region	Expect Value
serpin	1..136	58/142 (41%) 120/142 (85%)	1.3e-54
serpin	137..362	117/233 (50%) 208/233 (89%)	5.2e-115

5

### Example 8.

The NOV8 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 8A.

Table 8A. NOV8 Sequence Analysis			
	SEQ ID NO: 31	697 bp	
NOV8a, CG163133-02 DNA Sequence	CAGTGTGCTGGAATTCGCCCTTAACCGGCAGGATGTGGAGGTGCGGCTGCCACCGCTACGCGC CCTGGACGACTTTGTTCTGGGGTCGGCGCGTCTGGCGGCTCCGGATCCATGCGACCCGCGCGCA TGGTGCCACCGCGTCATCAACAACCTCCTCTACTACCAAACTACCTTCTCTGCTTCGGCA TCGGCCTCGCTCTCGCCGGGTACGTGCGGCCACTTCATACGCTCCTGAGCGCGCTGGTAGTGGC GGTGGCCCTCGGCGTGTGGTGTGGGCAGCTGAGACCCGCGCAGCTGTGCGCCGCTGCCGCCGC AGCCACCCTGCAGCCTGCCTGGCCGAGTGCTTGCCGTGCGGCTCCTGGTGTCTGGGTGCGCGG GCGGCGCTTGCACTTCCTGTTTCAGCATCGCCGGGCGGTGCTTCTGATCCTGGTGACAGCCTC GTTGCGCCTGCGCAACCTTAAGAACAGATTGAGAACAGATCGAGAGCATTGGTCTCAAGCGG ACGCCAATGGGCTGTACTAGAGGCACTGGGCAAGAGCAGGAGGCTGGATCCTAGGCCCTTG GGATCTGTACCCAGGACCTGGAGAAATACCAACCCACCCAGCCATAATTGGGACCCAGAGCC CTTTCCAGCACTTAAACAGGAGCCTAGAGCCGCTGCCCAACAAAAAGGGCGA		
	ORF Start: ATG at 33		ORF Stop: TAG at 567

10

	SEQ ID NO: 32	178 aa	MW at 19257.6kD
NOV8a, CG163133-02 Protein Sequence	MSEVRLPPLRALDDFVLGSARLAAPDPCDPQRWCHRVINLLYYQTNYLLCFGIGLALAGYVRP LHTLLSALVVAVALGVLVWAAETRAAVRRCRSHPAACLAAVLAVGLLVWVAGGACTFLFSIA GPVLLILVHASLRLRLNKNKIENKIESIGLKRTPMGLLLEALGQEQAQS		

	SEQ ID NO: 33	581 bp	
NOV8b, CG163133-01 DNA Sequence	GCAGTGTGTGGAATCGCCCTTAACCGGCAGGATGTCGGAGGTGCGGCTGCCACCGCTACGCGCC CTGGACGACTTGTTCCTGGGGTCGGCGCGTCTGGCGGCTCCGGATCCATGCGACCCGCGAGCGAT GGTGCCACCGCGTCATCAACAACCTCCTCTACTACCAAAACCACTACCTTCTCTGCTTCGGCAT CGGCCTCGCTCTCGCCGGGCAGTGCAGCCACTTCATACGCTCCTAAGCGCGCTGGTAGTGGCG GTGGCCCTCGGCGTCTGGTGTGGGCAGCTGAGACCCGCGCAGCTGTGCGCCGCTGCCGCCGCA GCCACCTGCGAGCTGCCTGGCGCGAGTGCCTTGGCGTCCGCTCCTGGTGACGCGCTCGTTGCG CCTGCGCAACCTTAAGAACAGATTGAGAACAGATCGAGAGCATTGGTCTCAAGCGGACGCCA ATGGGCCTGCTACTAGAGGCACTGGGACAAGAGCAGGAGGCTGGATCCTAGGCCCTGGGATCT GTACCCAGGACCTGGAGAATACCACCCACCCCCAGCCCATATTGGGACCCAGAGCCCTTTCC CAGCA		
	ORF Start: ATG at 32		ORF Stop: TAG at 497

	SEQ ID NO: 34	155 aa	MW at 16888.7kD
NOV8b, CG163133-01 Protein Sequence	MSEVRLPPLRALDDFVLGSARLAAPDPCDPQRWCHRVINNLYYQTNLLCFGIGLALAGHVRP LHTLLSALVVAVALGVLVWAAETRAAVRRCRRSHPAACLA AVLAVGLLVHASIRLRNLKNTIEN KIESIGLKRTPMGLLLEALGQEQEAGS		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 8B.

Table 8B. Comparison of NOV8a against NOV8b.		
Protein Sequence	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV8b	1..178 1..155	154/178 (86%) 155/178 (86%)

10

Further analysis of the NOV8a protein yielded the following properties shown in Table 8C.

Table 8C. Protein Sequence Properties NOV8a	
SignalP analysis:	No Known Signal Sequence Indicated
PSORT II analysis:	PSG: a new signal peptide prediction method N-region: length 10; pos.chg 2; neg.chg 1 H-region: length 2; peak value -2.04 PSG score: -6.44  GvH: von Heijne's method for signal seq. recognition GvH score (threshold: -2.1): -3.41 possible cleavage site: between 59 and 60  >>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation  
 Init position for calculation: 1  
 Tentative number of TMS(s) for the threshold 0.5:

4	INTEGRAL	Likelihood = -0.59	Transmembrane
49 - 65	INTEGRAL	Likelihood = -9.66	Transmembrane
68 - 84	INTEGRAL	Likelihood = -10.14	Transmembrane
100 - 116	INTEGRAL	Likelihood = -7.38	Transmembrane
120 - 136	PERIPHERAL	Likelihood = 8.43 (at 155)	

ALOM score: -10.14 (number of TMSs: 4)

MTOP: Prediction of membrane topology (Hartmann et al.)  
 Center position for calculation: 56  
 Charge difference: 1.0 C( 1.5) - N( 0.5)  
 C > N: C-terminal side will be inside

>>> membrane topology: type 3b

MITDISC: discrimination of mitochondrial targeting seq  
 R content: 2 Hyd Moment(75): 4.19  
 Hyd Moment(95): 1.05 G content: 0  
 D/E content: 2 S/T content: 1  
 Score: -5.78

Gavel: indication of cleavage sites for mitochondrial preseq  
 cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals  
 pat4: none  
 pat7: none  
 bipartite: none  
 content of basic residues: 10.1%  
 NLS Score: -0.47

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals:  
 XXRR-like motif in the N-terminus: SEVR

none

SKL: peroxisomal targeting signal in the C-terminus:  
 none

PTS2: 2nd peroxisomal targeting signal: none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:  
 type 1: none  
 type 2: none

NMYR: N-myristoylation pattern : none

	<p>Prenylation motif: none</p> <p>memYQRL: transport motif from cell surface to Golgi: none</p> <p>Tyrosines in the tail: none</p> <p>Dileucine motif in the tail: none</p> <p>checking 63 PROSITE DNA binding motifs: none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p> <p>checking 33 PROSITE prokaryotic DNA binding motifs: none</p> <p>NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination Indication: cytoplasmic Reliability: 94.1</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions total: 0 residues</p> <p>-----</p> <p>Final Results (k = 9/23):</p> <p>55.6 %: endoplasmic reticulum 11.1 %: mitochondrial 11.1 %: Golgi 11.1 %: vacuolar 11.1 %: cytoplasmic</p> <p>&gt;&gt; indication for CG163133-02 is end (k=9)</p>
--	---

A search of the NOV8a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 8D.

5

Table 8D. Geneseq Results for NOV8a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE14754	Human CCR5 chemokine receptor-interacting protein P2 - Homo sapiens, 178 aa. [EP1207202-A1, 22-MAY-2002]	1..178 1..178	178/178 (100%) 178/178 (100%)	2e-98
ABB97608				2e-98

	- Homo sapiens, 178 aa. [WO200222660-A2, 21-MAR-2002]	1..178	178/178 (100%)	
AAB94612	Human protein sequence SEQ ID NO:15456 - Homo sapiens, 178 aa. [EP1074617-A2, 07-FEB-2001]	1..178 1..178	177/178 (99%) 178/178 (99%)	4e-98
AAE14761	Human CCR5 chemokine receptor-interacting protein P2 mutant (G53A) - Homo sapiens, 178 aa. [EP1207202-A1, 22-MAY-2002]	1..178 1..178	177/178 (99%) 177/178 (99%)	1e-97
AAE14760	Human CCR5 chemokine receptor-interacting protein P2 mutant (G157R) - Homo sapiens, 178 aa. [EP1207202-A1, 22-MAY-2002]	1..178 1..178	177/178 (99%) 177/178 (99%)	2e-97

In a BLAST search of public sequence databases, the NOV8a protein was found to have homology to the proteins shown in the BLASTP data in Table 8E.

Table 8E. Public BLASTP Results for NOV8a				
Protein Accession Number	Protein/Organism/Length	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O60831	JM4 protein - Homo sapiens (Human), 178 aa.	1..178 1..178	178/178 (100%) 178/178 (100%)	5e-98
Q9JIG8	DXImx39e protein (DNA segment, Chr X, Immunex 39, expressed) - Mus musculus (Mouse), 178 aa.	1..178 1..178	162/178 (91%) 166/178 (93%)	6e-89
Q9ES40	Glutamate transporter EAAC1 interacting protein - Rattus norvegicus (Rat), 188 aa.	3..176 2..175	78/174 (44%) 119/174 (67%)	4e-41
O75915	JWA protein (HSPC127) (Vitamin A responsive, cytoskeleton related) - Homo sapiens (Human), 188 aa.	3..175 2..174	79/173 (45%) 117/173 (66%)	4e-41
Q9DB37	5930404D22Rik protein (RIKEN cDNA 5930404D22 gene) (JWA protein) - Mus musculus (Mouse), 188 aa.	3..175 2..174	78/173 (45%) 118/173 (68%)	1e-40

5

#### Example 9.

The NOV9 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 9A.

Table 9A. NOV9 Sequence Analysis			
	SEQ ID NO: 35	6240 bp	
NOV9a, CG165528-01 DNA Sequence	CTTTCTGTCTCTCGGGACCCCTTATTTCCTTCGTACGGTGTCCAGGACCATTGACCTGTCCGG CCCCAGGACCCCCCGCCGACCCAGCCCCGAGCATGGGGACGGCGCTGTCCAGCGCGGGGG CTGTTTTCTCTGTGCTCTCGCTGTCTCTGGGCTGTGGGCGGAGCTGGGCGAGCGGGCTG GAGTTTTCCGGGCGCCGAGGGGCAATGGACGCGCTTCCCCAAGTGGAACGCGCTGTGCGAGAGCG AGATGAGCTTCCAGCTCAAGACTCGCAGCGCCCGCGGCTCTGTCTTACTTCGACGACGAGGG CTTCTGCGACTTCTCGGAGCTGATTCTGACGCGCGGCGGCGCTGCAGCTCAGCTTCTCCATC TTCTGCGCTGAGCCTGCGACGCTCTGGCCGACACGCGGTAAACGACGGCGCTGGCAACGG TGCGCATCCGCGCCAGTTCCGCAACACACGCTCTTCATCGACAGGTGGAGGCCAAGTGGGT GGAGGTCAAGTCCAAGCGCAGGGACATGACGGTGTTCAGCGGCCCTTTTCGTGCGGGGGCTGCC CCGGAATGCGCGCCGCGGCGCTCAAGCTCACCTGGCCTCGGTGAGGGAGCGGAGCCCTTCA AGGGGTGGATTCTGTGACGTGAGGGTCAACTCCTCGCAGGTCTCGCCGTGGACAGCGCGAGGT GAAGCTGGACGATGAGCCGCCAACAGCGCGGGGGAAGCCCGTGGCAGGCGGGCGAGGAGGGC GAGGGCGGGGTGTGCTCAACGAGGGTGTGTCTCCGTGGTGGACGACAGGCGGTGTGCGACT GCTCGGAAACCGGCTTCCGCGCAAGGACTGCAGCCAAGAAGACAACTGTGGAAGGTCTGGC GCACCTGATGATGGGCGACCAAGGTAAAGTAAAGGAAAAGAAATATATTGCCACGTTCAAA GGATCTGAATACTTCTGTACGACTTGTCTCAAAACCCATTCAAGCAGCAGTGATGAAATAA CTCTGTCAATTAATAACCTTCAGAGGAATGGACTGATGCTTCACACTGGGAAATCGGCTGATTA TGTCAATCTTGCCCTGAAAAATGGAGCTGTCTCTGTGTCATTAATTTGGGATCAGGGGCTTT GAAGCACTAGTGGAGCCTGTGAATGGAAGTTAATGATAATGCCTGGCATGATGTGAAGTCA CCAGGAATCTGCGTCAGCACTCAGGCATTGGACACGCTATGGTAAACAACTACATTGTTCGGT GACAATATCAGTGGATGGGATTCTTACCACAACGGGCTACACGCAAGAAGATTATACCATGCTG GGGTCTGATGACTTTTTCTATGTTGGAGGCGAGTCCAGCACAGCCGACCTTCCAGGGTCAACG TCAGTAAACAATTTATGGGCTGTCTCAAGAGGTTGTATATAAAATAATGATGTGAGGCTGGA ATTATCTCGACTTGCCAGCAAGGAGATCCTAAGATGAAGATCCATGGAGTGGTGGCATTTAA TGTGAGAATGTTGCACTTTAGACCAATCACCTTTGAAACCCAGAGTCTTTTCATCTCTTTCG CTAAATGGAATGCAAGAAAACTGGCTCCATATCATTTGATTTCCTGTACACAGAGCCAAATGG CCTCATCTTATTAGCCATGGCAAGCCAAAGACATCAGAAAGATGCCAAGCACCCACAGATGATA AAGGTGCACTTCTTGTCTATTGAGATGCTAGATGGCCACCTCTACCTCCTCGGACATGGGGT CAGGTACTATAAAAAATAAAGCCCTGTTGAAGAAAGTGAATGATGAGAGATGATCATGTGGA CTTCAGAGAGACGGACGGTCAGGTACCATTTCTGTCAACAGTTGCGTACTCCCTACACTGCT CCTGGTGGAGTGGATTCTGGACCTGGATGATGAGTTGTACTCTGGGGGGGCTGCCAGAAAAATA AAGCTGGCCTTGTCTTCCCAACCGAGGTGTGGACTGCTCTGCTCAACTATGGCTACGTGGGCTG CATCAGGGATTGTTCTATCGATGGCCAAAGCAAGATATCCGCAAAATGGCTGAAGTTCAAAGT ACTGCTGGAGTGAAGCCTTCTGCTCAAAGGAAACAGCAAAACCGTGCCTTAGCAACCCCTGCA AAAACAATGGCATGTGACAGGATGGGTGGAACAGATATGCTGTGATTGTTCCGGAAACAGGCTA TCTTGGCAGGTCTGTGAGAGAGAGGCAACGGTTTTGAGCTATGATGGGAGCATGTTTATGAAA ATTACGCTCCCCGTAGTCATGCATACGAGGCTGAGGATGTTTCTTACGGTTCGGATCCAGC GTGCATATGGCATTCTGATGGCAACCACTTAGAGACTCTGTGACACCTCCGCTCGGAGCT AGACGCAGGACGTGTGAACTGACGGTCAATCTAGATTGTATCAGGATTAAGTGAATTCAGC AAAGGTCCCGAGACTCTTTTGTGCTGCTATAACCTCAATGATAACGAGTGGCACACAGTGCCTG TAGTTCCGCGTGGAAAAAGTTTAAAGTTAACAGTGGATGACCAACAGGCCATGACAGGTCAAAT GGCAGGTGATCATACTAGGCTGGAGTTCCATAACATAGAGACTGGCATCATCAGAACGACGG TATCTTTCTTCTGTCCCCTCCAACCTTCAATGGACACCTGCAGAGCTTGACATTTAATGGAATGG CATACATTGACCTGTGTAAATAATGGCGACATAGATTACTGTGAGCTTAATGCCAGATTGCGCTT CAGGAACATCATAGCAGATCTGTCACTTCAAGACCAATCGAGCTATGTGCTTAGCTACC TTGCAAGCCTACACTTCTATGCATCTTTTTTCCAGTTCAAGACCAACATCCCTAGATGGATTAA TTCTATATAACAGTGGGGATGGAATGACTTTATTGTGGTTGAATTAGTTAAAGGTACTTACA TTACGTGTTGATTGTTGGGAAATGGTGCTAACCTCATCAAAGGAAGCTCAAATAAACCCTCTCAAT GACAATCAGTGGCACACGTGATGATATCAAGGGACACAGCAACCTCCCACTGTAAAGATTG ACACAAAAATCAACACGCAATCACCGCCGGAGCCAGGAACCTTAGACCTCAAGAGTGACTTATA TATAGGAGGAGTAGCTAAAGAAACATACAAATCCTTACCAAACTCTGTACATGCCAAAGAGGC TTTCAAGGCTGCTGGCATCAGTTGATTAAATGGACGGCTTCCGACCTCATCTCCGATGCTC TTTTCTGCAACGGACAGATCGAGAGAGGATGTGAAGGGCCAGCAACCTGCCAAGAGGACTC ATGTTCCAATCAAGGTGTGTCTTGCAACAATGGGATGGCTTCAGCTGTGACTGTAGTATGACT TCCTTCAGTGGACCACTCTGCAATGACCCTGGGACGACATATATCTGTAGCAAGGTGGTGGAC AAATCAGTATAAGTGGCCTCTAATGACCGACCCAGTACACGAGCAGACAGACTGGCCATAGG TTTTAGCACTGTTTCAAGAAAGCCGTTATGGTGGAGTGGACAGTTCTTCAAGCTTGGGTGAC TACCTAGAATGCAATATACACAGGAAAAATGGAGTTAAGTTAATGTTGGACAGATGACA		

	TCGCCATTGAAGAATCCAATGCAATCATTAAATGATGGGAAATACCATGTAGTTCTGTTTCACGAG GAGTGGTGGCAATGCCACGTTGCGAGGTGGACAGCTGGCCAGTGATCGAGCGCTACCTGCAGGA AACAAATGATAACGAGCGCCTGGCGATTGCTAGACAGCGAATTCCATATCGACTTGGTCGAGTAG TTGATGAATGGCTACTCGACAAAGGGCGTCAGCTCAAACTCTTCAATAGCCAAGCAACCCATAAT AATTGGCGGAAAGACAGAGGCCAGCCCTTCCAGGGCCAGCTCTCTGGGCTGTACTACAATGGC TTGAAAGTTCTGAATATGGCAGCCGAAAACGATGCCAACATCGCCATAGTGGGAAATGTGAGAC TGGTTGGTGAAGTGCTTCTCTATGACAACTGAGTCAACAGCCACTGCCATGCAATCAGAGAT GTCCACATCAATTATGGAGACTACCACGACCCTGGCTACTAGCAAGCCAGAGGAGGAAAGCCC CCGACAAAAGAACCCATTAGCCAGACCACAGATGACATCCTTGTGGCCTCAGCAGAGTGTCCCA GCGATGATGAGGACATTGACCCCTGTGAGCCGAGCTCAGGTGGGTTAGCCAAACCCACCCGAGC AGGCGGCAGAGAGCCGTATCCAGGCTCAGCAGAAGTGATCCGGGAGTCCAGCAGACCACGGGT ATGGTCGTGGGATAGTAGCCGCTGCCGCCCTGTGCATCCTTATCCTCCTATGTCCTATGTACA AGTACAGAAACCGGGATGAAGGCTCATACCATGTGGACGAGAGTCGAAACTACATCAGTAACCTC AGCACAGTCCAATGGGGCTGTTGTAAAGGAGAAACAACCCAGCAGTGCGAAAAGCTCCAACAAA AATAAGAAAAACAAGGATAAAGAGTATTATGTCTGATCCCAAGATCTTAAATGGACACTTGTAT AGAAATAGTCTTCAATTTATCTGAGACATAATATAAACTTATTTACTTTCTTTTATGAAGCA CATACAAAAGAACAGAGAATGCAATCAGGAAGGAAAGACTTTTAAAAATAAAAAACAAGTA TCTCATGCTCTTGTCTTCTCAAAAAAGAAAAACAAAAACAAAAACAGGGGCCAATAAATCCC TAACATCCACAGTGTTTTCATTTACTCTGCTTGTCTTTATGTTGCTGGAACTTTCTAAAAGAC AGTGATGACCGCAGCATTCTAAAAGCAAAGGAGTACTACAGCATCAAGGCACAACAACAAAAAC CAACACAAAAACAATAACAAAAAGAGCTACCTATGATCCTGGATTAGCCAAAGTGTAGCG CTTTCCTGAGAAGTCAGTCCAATTGCCAGAGAAGACTGTCTTTTGAAGTACTCAACCTGCAAA CCTTTAAGAGTTTGGCCGCTGGTGCACCTGGAGCAGTGGTTGGAACCTTGCAATTGAAACAAAGT GCTGGCTTTTTTGAAGACTTGTGTAGGAACACTTCAAAAAGCCCTTTCTGGTTGTGAGAGAG GAAAAAAAAGTATGGAGGCCCTTATTTTCAAAAATGTGAAATATAAGGCAGCTTTTCACAAAA ATTTCAAAAACAAAAACAAGAGGGCATAGATGCAATCATTGGGAAATTTTATGCACGCTTATTA TGTTATTACATATGTTTATATAAAATCCATCTCTGTGTGCTTTCTGGACTGTGATAAGTGACGT TTTATAGCCTGTGTATAGAAAATGCAAAATATATCTCTGCTCTTCAGCCATTTTGGTAAAT CAATGTTATAAGTGTGTGCTAAGTATAGGGAGTTTTATGACATCAGAGCAACAATTATTTTCAGTT GGGTTTTTCTTTTTTTGGCCACATTATAAATTGCCACAATTACTTACTTTTATTTTTTAAAGA AATTACAGTGTAGTGTATTATCTTAAGGAAGATATGTATGAATGTATATACAAGAGCTCAGCTAC TTCTTTCTTATATGTAAGCCTTATTCTGTTGCAATTAAGTTTATAGTACTGTATGAAAGGT GTGAATTAGAAAGTCACATATATACATATGTATCTTATAATCTTTTCTCCCTGAAATACTCACA TTCCACATACATTTCACTATTTTACACACACACACACACACACACACACACACACACACAC ACGAATCCACAGCAATCCATCAGATATGCTGGAAGATCCAAACGTCATACAGTAGCAAAATT TATTGACAAATTGAAAAGCAGGAAGGAAGAGGGTTGTGCCAAGGTATTGATGACAAATGGGGT ATTTGCTTCATTGAGATCTTGCTCCAGGTAACCTTAAAGAGATTTTAGTCCCTAAAGAAATGA ACCTTTCTTATCAAATAGAATATCACTGATATCTGCTGCATGAATAAGAACCATATGTGGG CAGGTTATGGAAGCAAAATTGGTTAATCTACACCTTAACTCTGGCTGCTGCAATTGAAAACCTT CTTTCTAATAAAATAATATATATATCTCTGAA	
	ORF Start: ATG at 100	ORF Stop: TGA at 4642

	SEQ ID NO: 36	1514 aa	MW at 166226.0kD
NOV9a, CG165528-01 Protein Sequence	MGTALLQRGGCFLLCLSLILLGCWAEGLGSGLEFPFPAEGQWTRFPKWNACCSEMSFQLKTR SAR GLVLYFDDEGFCDLELILTRGGRLQLSFSIFCAEPATLLADTPVNDGAWHSVRIRQRFRNTTL FIDQVEAKWVEVKSRRDMTVFSGLFVGGLPPELRAAALKLTLASVREREPFGKWIRDVRVNSS QVLPVDSGEVKLDDPEPNSGGGSPCEAGEEGEGGVCLNGGVCSVVDDQAVCDCSRGTGFRGKDCS QEDNNVEGLAHLMMGDQKSKGKEEYIATFKGSEYFCYDLSQNPIQSSSDEITLSPKTLQRNGL MLHTGKSADYVNLALKNGAVSLVINLGSAGFALEVPVNGKFNDNAHWDVKVTRNLRQHSGIGH AMVKNLHCSVTISVDGILTTGYTQEDYTMGLSDDFFVYVGGSPSTADLPGPSVSNFMGCLKEV VYKNNVRLLESLRLAKQGDPMKIHGVVAFKCNVATLDPITFETPESFISLPKWNAKKTGSIS FDFRTTEPNGLILFSGHKPRHQDAKHQPMIKVDFFAIEMLDGHLYLILDMGSGTIIKALKKK VNDGEWYHVDQDRSGTISVNTLRTPYTAPGESEILDLDDELGLGLPENKAGLVFPTEVWT ALLNYGVGCIKRDIFIDGQSKDIQMAEVQSTAGVKPSCSKETAKCLSNPKNNNGMCRDGNWR YVCDSCSGTGYLGRSCEREATVLSYDGSFMFKIQLPVVMHTEAEDVSLRFRSQRAYGILMATTSR DSADTLRLLEDAGRVLKLTNLDICIRINCNSKGPETLFAGYNLNDNEWHTVRRVRRGSKSLTLV DDQQAMTQMGADHTRLEPHNIETGIIITERRYLSSVPSNFIHGLQSLTFNGMAYIDLCKNGDID YCELNARFGFRNIADPVTFTKSSVVALATLQAYTSMHLFFQFKTSLDGLILYNSGDGNDPI VVELVKGYLHYVFDLNGANLIKSSNKPLNDNQWHNVMSRDTSNLHTVKIDTKITQITAGA RNLDLKSPLYIGGVAKEYSLPLKLVHAKGEGFQGLASVDLNGRLPDILSDALFCNQGIERGCE GPSTTCQEDSCSNQGVCLQQWDFGSCDCSMTSFGSGLCNDPGTTYIFSKGGQITYKWPPNDRP STRADRLAIGFSTVQKEAVLVRVDSGSLGDLYLELHIGQKIGVKFNVGTDDIAEESNAIIND GKYHVVRPTRSGGNATLQVDSWPVIERYPAGNNDNERLAIRQRIPLGRVVDWLLDKGRQL TIFNSQATIIIGGKEQGQPFQGLSGLYNGLVNMAAENDANIAIVGNVRLVGEVPSSMTTE		

STATAMQSEMSTSIMETTTTLATSTARRGKPPTKEPISQTTDDILVASAECPSDDEDIDPCEPS SGGLANPTRAGGREPYPGSAEVIRESSTTGMVVGIVAAAAALCILILLYAMYKYNRNRDEGSYHV DESRNYISNSAQSGAVVKEQPSSAKSSNKNKNKDKKEYVV
---

	SEQ ID NO: 37	1611 bp	
NOV9b, CG165528-02 DNA Sequence	AAACTTTGCTCTCCGCGGGCGGCTGCCCCCTCGGCGGGCGCCCCGCCATGTACCAGAGGATGCTCC GGTGC GGCGCCGAGCTGGGCTCGCCCGGGGGCGCGCGGGCGGGCGGGCGGGCGGCAGG GGGGCGCCTGGCCCTGCTTTGGATAGTCCCGCTCACCCCTCAGCGGCCTCTAGGAGTGGCGTGG GGGGCATCCAGTTTGGGAGCGCACCACATCCACCATTTCCATGGCAGCAGCAAGCATCATTTCAG TGCCTATTGCAATCTACAGGTCAACGGCATCCTTGCAGCGCCGACACGCTGGGACGACATATAT CTTTAGCAAAGGTGTGGACAAATCAGTATAAGTGGCTCTTAATGACCGACCCAGTACACGA GCAGACAGACTGGCCATAGGTTTTAGCACTGTTCAGAAAGAAGCCGTATTGGTGCAGTGGACA GTTCTTCAGGCTTGGGTGACTACCTAGAACTGCATATACACCAGGGAAAAATTGGAGTTAAGTT TAATGTTGGGACAGATGACATCGCCATTGAAGAACTCCAATGCAATCATTAAATGATGGGAAATAC CATGTAGTTCGTTTTACGAGGAGTGGTGGCAATGCCACGTTGTCAGGTGGTACAGCTGGCCAGTGA TCGAGCGCTACCCCTGCAGGAAACAATGATAACGAGCGCCTGGCGATTGCTAGACAGCGAATTTC ATATCGACTTGGTCGAGTAGTTGATGAATGGCTACTCGACAAAGGGCGTCAGCTCAACATCTTC AATAGCCAAAGCAACCATATAATTTGGCGGGAAGAGCAGGGCCAGCCCTTCAGGGCCAGCTCT CTGGGCTGTACTACATAGGCTTGAAAGTTCTGAATATGGCAGCCGAAACGATGCCAACATCGC CATAGTGGGAAATGTGAGACTGGTTGGTGAAGTGCCTTCCTCTATGACAACTGAGTCAACAGCC ACTGCCATGCAATCAGAGATGTCCACATCAATTATGGAGACTACCACGACCTCGCTACTAGCA CAGCCACAAGAGAGAAAGCCCCGACAAAGAACCATTAGCCAGACCAAGATGACATCTCTGT GGCCTCAGCAGAGTGTCCAGCGATGTAGGACATTGACCCCTGTGAGCCGAGCTCAGTGGGG TTAGCCAACCAACCCGAGCAGGCGGCAGAGAGCCGTATCCAGGCTCAGCAGAAGTGATCCGGG AGTCCAGCAGCACCAAGGATATGGTTCGTTGGGATAGTAGCGCGCTCGCCGCTGTGCATCTCTAT CCTCCTCTATGCCATGTACAAGTACAGAAACGGGATGAAGGCTCATACCATGTGGACGAGAGT CGAAACTACATCAGTAACTCAGCACAGTCCAATGGGGCTGTTGTAAGGAGAAACAACCCAGCA GTGCGAAAAGCTCCAACAATAAAGAAAAACAAGGATAAAGAGTATTATGTCTGATCCCAAGA TCTTAAATGGACACTTGTATAGAAATAGTCTTTCATTTTATCTGAGACATAAATATAAACTTATTT ACTTTCCTTTTATGAAGCACATACAAAAGAAGACAGGGAATGCAATCAGGAAGGAAAGACTTT TTAAAAAATAA		
	ORF Start: ATG at 46		ORF Stop: TGA at 1462

5

	SEQ ID NO: 38	472 aa	MW at 50423.1kD
NOV9b, CG165528-02 Protein Sequence	MYQRLRCGAELGSPGGGGGGGGGGAGGRLALLWIVPLTSLGLGVANGASSLGAHHIHHFHG SSKHHSVPIAIYRSPASLRGGHAGTTTYIFSKGGGQITYKWPPNDRPSTRADRLAIGFSTVQKEA VLVRVDDSSSLGDDYLELHHQKGIQVKFNVGTDDIAIEESNAINDGKYHVVRPTRSGGNATLQ VDSWPVIERYPAGNVDNLERLAIARQIPYRLGRVVDVEWLLDKGRQLTIFNSQATIIGGKEQG PFQQLSLGLYINGLKVLNMAENDANIAIVGNRLVGEVLPSSMTTESATAMQSEMSTINMETT TTLATSTARGKPPPTKEPISQTTDDILVASAECPSDDDEDIDCEPSSGGLANPFRAGGREPYPG SAEVIRESSSTTGMVVGIVAAALCILILLIYAMYKYRNRDEGSYHVDESRNYISNSAQSGNAV KEQPSSAKSSNNKKNKDKEYV		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 9B.

10

Table 9B. Comparison of NOV9a against NOV9b.		
Protein Sequence	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV9b	1130..1514 88..472	385/385 (100%) 385/385 (100%)

Further analysis of the NOV9a protein yielded the following properties shown in Table 9C.

5

Table 9C. Protein Sequence Properties NOV9a	
SignalP analysis:	Cleavage site between residues 26 and 27
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 8; pos.chg 1; neg.chg 0  H-region: length 17; peak value 10.61  PSG score: 6.21</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): 7.83  possible cleavage site: between 25 and 26</p> <p>&gt;&gt;&gt; Seems to have a cleavable signal peptide (1 to 25)</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 26  Tentative number of TMS(s) for the threshold 0.5:  2  Number of TMS(s) for threshold 0.5: 1  INTEGRAL Likelihood = -13.59 Transmembrane  1440 -1456  PERIPHERAL Likelihood = 2.44 (at 89)  ALOM score: -13.59 (number of TMSs: 1)</p> <p>MTOP: Prediction of membrane topology (Hartmann et al.)  Center position for calculation: 12  Charge difference: -5.0 C(-3.0) - N( 2.0)  N &gt;= C: N-terminal side will be inside</p> <p>&gt;&gt;&gt; membrane topology: type 1a (cytoplasmic tail 1457 to 1514)</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 1 Hyd Moment(75): 10.18  Hyd Moment(95): 7.78 G content: 4  D/E content: 1 S/T content: 2  Score: -4.63</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  R-2 motif at 18 QRG GC</p>

	<p>NUCDISC: discrimination of nuclear localization signals  pat4: none  pat7: none  bipartite: none  content of basic residues: 10.4%  NLS Score: -0.47</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals:  KKXX-like motif in the C-terminus: KEYY</p> <p>SKL: peroxisomal targeting signal in the C-terminus:  none</p> <p>PTS2: 2nd peroxisomal targeting signal: none</p> <p>VAC: possible vacuolar targeting motif: none</p> <p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif:  type 1: none  type 2: none</p> <p>NMYR: N-myristoylation pattern : none</p> <p>Prenylation motif: none</p> <p>memYQRL: transport motif from cell surface to Golgi:  none</p> <p>Tyrosines in the tail: too long tail</p> <p>Dileucine motif in the tail: none</p> <p>checking 63 PROSITE DNA binding motifs:  Leucine zipper pattern (PS00029): *** found ***  LQRGGCFLLCLSLLLLGCWAEL at 6  none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p> <p>checking 33 PROSITE prokaryotic DNA binding motifs:  none</p> <p>NNCN: Reinhardt's method for Cytoplasmic/Nuclear  discrimination  Indication: cytoplasmic  Reliability: 70.6</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions  total: 0 residues</p> <p>-----</p> <p>Final Results (k = 9/23):</p> <p>44.4 %: endoplasmic reticulum  22.2 %: Golgi</p>
--	---

	11.1 %: plasma membrane 11.1 %: vesicles of secretory system 11.1 %: extracellular, including cell wall  >> indication for CG165528-01 is end (k=9)
--	---

A search of the NOV9a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 9D.

5

<b>Table 9D. Geneseq Results for NOV9a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV9a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAM79855	Human protein SEQ ID NO 3501 - Homo sapiens, 1522 aa. [WO200157190-A2, 09-AUG-2001]	1..1514 47..1522	1465/1517 (96%) 1466/1517 (96%)	0.0
AAM78871	Human protein SEQ ID NO 1533 - Homo sapiens, 1327 aa. [WO200157190-A2, 09-AUG-2001]	147..1514 1..1327	1326/1368 (96%) 1326/1368 (96%)	0.0
AAE17600	Human extracellular messenger (XMES)-2 protein - Homo sapiens, 1438 aa. [WO200194587-A2, 13-DEC-2001]	19..1514 16..1438	1041/1496 (69%) 1210/1496 (80%)	0.0
AAU28190	Novel human secretory protein, Seq ID No 359 - Homo sapiens, 1712 aa. [WO200166689-A2, 13-SEP-2001]	16..1411 14..1419	975/1426 (68%) 1162/1426 (81%)	0.0
AAU14241	Human novel protein #112 - Homo sapiens, 1091 aa. [WO200155437-A2, 02-AUG-2001]	414..1514 1..1091	834/1101 (75%) 960/1101 (86%)	0.0

In a BLAST search of public sequence databases, the NOV9a protein was found to have homology to the proteins shown in the BLASTP data in Table 9E.

10

Table 9E. Public BLASTP Results for NOV9a				
Protein Accession Number	Protein/Organism/Length	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q63372	Neurexin 1-alpha precursor (Neurexin I-alpha) - Rattus norvegicus (Rat), 1514 aa.	1..1514 1..1514	1496/1514 (98%) 1506/1514 (98%)	0.0
Q28146	Neurexin 1-alpha precursor (Neurexin I-alpha) - Bos taurus (Bovine), 1530 aa.	1..1514 1..1530	1503/1530 (98%) 1508/1530 (98%)	0.0
A40228	neurexin I-alpha precursor - rat, 1507 aa.	1..1514 1..1507	1489/1514 (98%) 1499/1514 (98%)	0.0
BAA25504	KIAA0578 protein - Homo sapiens (Human), 1542 aa (fragment).	1..1514 47..1542	1496/1514 (98%) 1496/1514 (98%)	0.0
BAC41433	MKIAA0578 protein - Mus musculus (Mouse), 1525 aa (fragment).	1..1514 47..1525	1468/1514 (96%) 1473/1514 (96%)	0.0

PFam analysis indicates that the NOV9a protein contains the domains shown in the Table 9F.

5

Table 9F. Domain Analysis of NOV9a			
Pfam Domain	NOV9a Match Region	Identities/ Similarities for the Matched Region	Expect Value
laminin_G	58..195	46/167 (28%) 101/167 (60%)	4e-12
laminin_G	312..378	23/81 (28%) 47/81 (58%)	1.4e-08
laminin_G	393..456	17/81 (21%) 43/81 (53%)	0.013
laminin_G	515..662	56/169 (33%) 114/169 (67%)	1.1e-28
EGF	687..719	13/47 (28%) 24/47 (51%)	0.00049
laminin_G	753..834	26/97 (27%) 59/97 (61%)	0.00016
laminin_G	940..1071	43/163 (26%) 99/163 (61%)	6.4e-16

EGF	1094..1126	12/47 (26%) 26/47 (55%)	0.00019
laminin_G	1163..1236	22/87 (25%) 48/87 (55%)	3.2e-07

**Example 10.**

The NOV10 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 10A.

5

Table 10A. NOV10 Sequence Analysis			
	SEQ ID NO: 39	1365 bp	
NOV10a, CG165666-01 DNA Sequence	ACGCGTGGAGTCCTGCGGGCCGTGGCCACCCAGCAGCGCGGCGCCGTGTCGTGGACAAGGAGA ACCTCACCATGCCGGGCCTCAGGTTGACAAACATCCAGGGAGATGCAGTTAAAGACTTGATGCT TCGCTTTCTGGGTGAAAAAGCTGCAGCAAAGAGACAAGTCCTAAATGCCGACTCAGTGAACAA TCTTTTGTGGATTGAAACAGCTAATCCTCTGGTTTGTGAGGCTGGCACTACTAGTGAAGTTGG GCCTTTTCCAGAATGCTGAGATGGAATTTGAACCCCTTCGGAAATCTTGATCAGCCAGATCTTTA TTACGAGTACTACCCGACGCTGTACCTGGGCGCAGGGGCTCCATGGTCCCCTTCTCGATGCGC ATCTTGACGCGGAGCTTCAGCAGTACCTGGGGAACCCACAGGAGTCGCTGGATAGACTGCACA AGGTGAAGACTGTCTGCAGCAAGATCCTGGCCAATTTGGAGCAAGGCTTAGCAGAAGACGGCGG CATGAGCAGCGTGACTCAGGAGGGCAGACAAGCCTCTATCCGGCTGTGGAGGTACGCTCTGGGC CGGGTGATGTACTCCATGGCAACTGTCTGCTCCTGATGAAGGATTATGTGCTGGCCGTGGAGG CGTATCATTTCGGTTATCAAGTATTACCCAGAGCAAGAGCCCCAGCTGCTCAGCGGCATCGGCCG GATTTCCCTGCAGATTGGAGACATAAAACAGCTGAAAAGTATTTTCAAGCGTTGAGAAAGTA ACACAGAAATTAGATGGACTACAGGTTAAATCATGGTTTGTGTAACAGCGCGTTCTTCCAC TCGGGCAGAATAACTTTGCAGAAGCCACAGGTTCTTCACAGAGATCTTAAGGATGGATCCAAG AAACGCAGTGGCCAAACAACAGCTGCCGTGTGTCTGCTCTACCTGGGCAAGCTCAAGGACTCC CTGCGGCAGCTGGAGGCCATGGTCCAGCAGGACCCAGGCACTACCTGCACGAGAGCGTGTCT TCAACCTGACCACTATGTACGAGCTGGAGTCTCAGGAGCATGCAGAAAGAACGCCCTTCTCT GGAGGCTGTGCGCGGCAAGGAGGGGACAGCTTCAACACACAGTGCCTCAAGCTGGCTTAGCTG CCTCCAACACACTACGTGAGAAGGACCCGGGTCTTTGAACTGTCTTGAAGCTAATGTATTA ATGTGACATGGAGGAATCAATAAACTCCTGCTTCACTGGTGTCTGCTGGTGTCTTCTTGGT CCCAAGCCACGGCCAGCCAGGACTTCCCGCAGTTGGTTCGGCGTTCAGCCACGCAGTCCCTG CAGCTGGGTCACTGTTTCATGA		
	ORF Start: ATG at 73		ORF Stop: TAG at 1147

	SEQ ID NO: 40	358 aa	MW at 40766.8kD
NOV10a, CG165666-01 Protein Sequence	MPGLRFDNIQGDVAVKDLMLRFLGEKAAAKRQVLNADSVESQSFVGLKQLILWFVRLALLVKLG LGFQNAEMEFEPFGNLDQPDLYEYYPHVYPGRRGSMVFPFMRILHAELOQYLGNPQESLDR LHKVKTVC SKILANLEQGLAEDGGMSSVTQBRQASIRLWRSRLGRVMYSMANCLLLMKDY VLAVEAYH SVIKYYPEQEPQLLSGIGRISLQIGDIKTAEKYFQDVEKVTQKLDGLQKIM VLMNSAFHLGQ NNFAEAHRFFTEILRMDPRNAVANNNAAVCLLYLGKLDKSLRQLEAMV QDPRHYLHESVLFNL TMYELESSRSMQKQALLEAVAGKEGDSFNTQCLKLA		

10 Further analysis of the NOV10a protein yielded the following properties shown in Table 10B.

Table 10B. Protein Sequence Properties NOV10a

SignalP analysis:	Cleavage site between residues 65 and 66
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 7; pos.chg 1; neg.chg 1  H-region: length 4; peak value -9.72  PSG score: -14.12</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): -6.97  possible cleavage site: between 58 and 59</p> <p>&gt;&gt;&gt; Seems to have no N-terminal signal peptide</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 1  Tentative number of TMS(s) for the threshold 0.5:  1  Number of TMS(s) for threshold 0.5: 1  INTEGRAL Likelihood = -5.47 Transmembrane  48 - 64  PERIPHERAL Likelihood = 3.87 (at 174)  ALOM score: -5.47 (number of TMSs: 1)</p> <p>MTOP: Prediction of membrane topology (Hartmann et al.)  Center position for calculation: 55  Charge difference: -4.0 C(-2.0) - N( 2.0)  N &gt;= C: N-terminal side will be inside</p> <p>&gt;&gt;&gt; membrane topology: type 2 (cytoplasmic tail 1 to 48)</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 1 Hyd Moment(75): 1.36  Hyd Moment(95): 3.47 G content: 2  D/E content: 2 S/T content: 0  Score: -8.05</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  R-2 motif at 15 LRF DN</p> <p>NUCDISC: discrimination of nuclear localization signals  pat4: none  pat7: none  bipartite: none  content of basic residues: 11.5%  NLS Score: -0.47</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals:  XXRR-like motif in the N-terminus: PGLR  KKXX-like motif in the C-terminus: CLKL</p> <p>SKL: peroxisomal targeting signal in the C-terminus:  none</p> <p>PTS2: 2nd peroxisomal targeting signal: found  KLKDSLRLQL at 292</p> <p>VAC: possible vacuolar targeting motif: none</p>

	RNA-binding motif: none
	Actinin-type actin-binding motif: type 1: none type 2: none
	NMYR: N-myristoylation pattern : none
	Prenylation motif: none
	memYQRL: transport motif from cell surface to Golgi: none
	Tyrosines in the tail: none
	Dileucine motif in the tail: none
	checking 63 PROSITE DNA binding motifs: none
	checking 71 PROSITE ribosomal protein motifs: none
	checking 33 PROSITE prokaryotic DNA binding motifs: none
	NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination Indication: cytoplasmic Reliability: 89
	COIL: Lupas's algorithm to detect coiled-coil regions
	218 I 0.66
	219 K 0.67
	220 T 0.71
	221 A 0.71
	222 E 0.71
	223 K 0.71
	224 Y 0.71
	225 F 0.71
	226 Q 0.71
	227 D 0.71
	228 V 0.71
	229 E 0.71
	230 K 0.71
	231 V 0.71
	232 T 0.71
	233 Q 0.71
	234 K 0.71
	235 L 0.71
	236 D 0.71
	237 G 0.71
	238 L 0.71
	239 Q 0.71
	240 G 0.71
	241 K 0.71
	242 I 0.71
	243 M 0.71
	244 V 0.71
	245 L 0.71
	246 M 0.71

247 N 0.71
248 S 0.71
249 A 0.71
total: 32 residues
-----
Final Results (k = 9/23):
39.1 %: mitochondrial
30.4 %: cytoplasmic
8.7 %: Golgi
8.7 %: nuclear
8.7 %: endoplasmic reticulum
4.3 %: vacuolar
>> indication for CG165666-01 is mit (k=23)

A search of the NOV10a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 10C.

5

Table 10C. Geneseq Results for NOV10a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB42120	Human ORFX ORF1884 polypeptide sequence SEQ ID NO:3768 - Homo sapiens, 379 aa. [WO200058473-A2, 05-OCT-2000]	18..358 1..379	341/379 (89%) 341/379 (89%)	0.0
ABB90440	Human polypeptide SEQ ID NO 2816 - Homo sapiens, 449 aa. [WO200190304-A2, 29-NOV-2001]	1..347 41..425	329/385 (85%) 335/385 (86%)	0.0
ABP61860	Human polypeptide SEQ ID NO 214 - Homo sapiens, 271 aa. [US2002065394-A1, 30-MAY-2002]	96..358 9..271	263/263 (100%) 263/263 (100%)	e-148
AAW73629	Human secreted protein clone cd265_11 - Homo sapiens, 271 aa. [WO9855614-A2, 10-DEC-1998]	96..358 9..271	263/263 (100%) 263/263 (100%)	e-148
ABB65708	Drosophila melanogaster polypeptide SEQ ID NO 23916 - Drosophila melanogaster, 484 aa. [WO200171042-A2, 27-SEP-2001]	1..342 95..463	126/386 (32%) 185/386 (47%)	4e-48

In a BLAST search of public sequence databases, the NOV10a protein was found to have homology to the proteins shown in the BLASTP data in Table 10D.

Table 10D. Public BLASTP Results for NOV10a				
Protein Accession Number	Protein/Organism/Length	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q8WVT3	Similar to TPR-containing protein - Homo sapiens (Human), 735 aa.	1..358 340..735	357/396 (90%) 358/396 (90%)	0.0
Q8K2L8	Hypothetical protein - Mus musculus (Mouse), 797 aa.	1..358 402..797	340/396 (85%) 351/396 (87%)	0.0
Q8WVW1	CGI-87 protein - Homo sapiens (Human), 379 aa.	18..358 1..379	341/379 (89%) 341/379 (89%)	0.0
Q9Y395	CGI-87 protein - Homo sapiens (Human), 379 aa.	18..358 1..379	339/379 (89%) 340/379 (89%)	0.0
Q8N9N0	Hypothetical protein FLJ36862 - Homo sapiens (Human), 696 aa.	1..278 323..644	276/322 (85%) 277/322 (85%)	e-149

- 5 Pfam analysis indicates that the NOV10a protein contains the domains shown in the Table 10E.

Table 10E. Domain Analysis of NOV10a			
Pfam Domain	NOV10a Match Region	Identities/ Similarities for the Matched Region	Expect Value
TPR	168..201	8/34 (24%) 23/34 (68%)	0.0053

#### Example 11.

- 10 The NOV11 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 11A.

Table 11A. NOV11 Sequence Analysis			
	SEQ ID NO: 41	3462 bp	
NOV11a, CG165676-01 DNA	GATGGGGCCAGAACGGACAGGGGCCGCGCTGCCGCTGCTGCTGGTGTAGCGCTCAGTCAA GGCATTTTAAATTGTTGTTTGGCCTACAATGTTGGTCTCCAGAAGCAAAAATATTTCCGGTC		

Sequence	CTTCAAGTGAACAGTTTGGCTATGCAGTGCAGCAGTTTATAAATCCAAAAGGCAACTGGTTACT GGTTGGTTTCAACCCTGGAGTGGCTTTCTGTGAGAACCGAATGGGAGATGTGTATAAATGTCTGTT GACCTATCCACTGCCACATGTGAAAACTAAATTTGCAAACTTCAACAAGCATTCCAATGTTA CTGAGATGAAAAACCAACATGAGCCTCGGCTTGATCCTCACCAGGAACATGGGAACATGGAGGTTT TCTCACATGTGGTCTCTGTGGGCACAGCAATGTGGGAATCAGTATTACACAACGGGTGTGTGT TCTGACATCAGTCTGTATTTTCACTCTCAGCCAGCTTCTCACCCTGCAACTCAGCCCTGCCCTT CCCTCATAGATGTTGTGGTTGTGTGTATGAATCAAATAGTATTTATCCTTGGGATGCAGTAAA GAATTTTTTGGAAAAATTTGTACAAGGCCTGGATATAGGCCCCACAAGACACAGGTGGGTTA ATTAGTATGCCAATAATCCAAGAGTTGTGTTTAACTTGAAACATATAAAACCAAGAGAGAAA TGATTGTAGCAACATCCAGACATCCCAATATGGTGGGGACCTCACAACACATCTCGGAGCAAT TCAATATGCAAGAAAATATGCTTATTTCAGCAGCTTCTGGTGGGCGACGAAGTGTACGAAAGTA ATGGTAGTTGTAACGTGACGGTGAATCAGATGATGGTTCAATGTTGAAAGCTGTGATTGATCAAT GCAACCATGACAATATACTAGGTTTGGCATAGCAGTTCTTGGGTGACTTAAACAGAAACGCCCT TGATACTAAAAATTTAATAAAAGAAATAAAAGCAATCGCTAGTATTCCAACAGAAAGATACTTT TTCAATGTGTCTGATGAAGCAGCTCTACTAGAAAAGGCTGGGACATAGGAGAAACAAATTTTCA GCATTGAAGGTACTGTTCAAGGAGGAGACAACCTTCAGATGGAATGTCAAGTGGGATTTCAG TGCAGATTACTCTTCTCAAAATGATATTCTGATGCTGGGTGACAGTGGGAGCTTTTGGCTGGAGT GGGACCATGTCCAGAAGACATCTCATGGCCATTGATCTTTCTTAAACAAGCCTTTGACCAAAA TTCTGCAGGACAGAAATCAGTTTCATATTTAGGTTACTCTGTGGCTGCAATTTCTACTGGAGA AAGCACTCACTTTGTGCTGGTGTCTCTCGGGCAAATTATACCGCCAGATAGTGTCTATATAGT GTGAATGAGAATGGCAATATCAGGTTTATTTCAGGCTCACCGAGGTGACAGATTGGCTCCTATT TTGGTAGTGTGCTGTGTTCACTTGATGTGGATAAAGACACCATTACAGACGTGCTCTTGGTAGG TGCAACCAATGTACATGAGTGACCTAAGAAAGAGGAAGGAGATCTACCTGTTTACTATCAAAA GAGGGCATTTTGGGTGAGCACCAATTTCTTGAAGGCCCGAGGCAATTGAAAACACTCGATTTG GTTTCAGCAATTGCAGCTCTTTTCAGACATCAACATGGATGGCTTTAATGATGTGATTGTTGGTTC ACCACTAGAAAATCAGAATTTCTGGAGCTGTATACATTTACAATGGTCATCAGGACATATCCGC ACAAAGTATTTCCAGAAAATCTTGGGATCCGATGGAGCCTTTAGGAGCCATCTCAGTACTTTG GGAGGTCTTGGATGGCTATGGAGATTAAATGGGGATTCCATCAGGATGTGCTATTGGTGC CTTGGACAAGTGGTCAACTCTGGTCACAAAGTATTGCTGATGTAGCTATAGAAGCTTCATTTC ACACCAGAAAAAATCACTTTGGTCAACAAGAAATGCTCAGATAATTCTCAAACCTCTGCTTCAGTG CAAAGTTCAGACCTACTAAGCAAAACAAATCAAGTGGCCATTGTATATAACATCACACTTGATGC AGATGGATTTTCATCCAGAGTAACCTCCAGGGGGTTATTAAAGAAAAACAAATGAAAGGTGCTG CAGAAGAAATATGGTAGTAATCAAGCACAGAGTTGCCCGGAGCACATCATTATATACAGGAGC CCTCTGATGTTGTCAACTCTTTGGATTGCGGTGTTGGACATCAGTCTGGAAGAACCTTGGCACTAG CCCTGCCCTTGAAGCCTATTCTGAGACTGCCAAGGTCTTCAGTATTCCTTTTCCAACAAGACTGT GGTGAGGACGGACTTTGCATTTCTGATCTAGTCTAGATGTCCGACAAATACCAGCTGCTCAAG AACCAACCTTTATTGTGACCAACCAAAACAAAGGTTAATCAATTTTCAGTAACGCTGAAAAATAA AAGGGAAGTGACATCAACTGGAATTTGTTGTTGATTTTTCAGAAAATTTGTTTGTGATCA TTCTCCCTGCCGGTTGATGGGACAGAAATCAATGCCAGGTGGCTGCATCTCAGAAGTCTGTTG CCTGCGATGTAGGCTACCTGCTTTAAAGAGAGAAACAACAGGTGACTTTTACTATTAACTTTGA CTTCAATCTTCAAACCTTCAGAAATCAGGCGTCTCTCAGTTTCCAGGCCTTAAGTGAAGCCAA GAAGAAAAACAAGGCTGATAATTTGGTCAACCTCAAAATTCCTCTCTGTATGATGCTGAAATTC ACTTAACAAAGGTAACAACAGGAAGTGTCCAGTAAGCATGGCAACTGTAATCATCCACATCCC TCAGTATACCAAGAAAAGAACCCACTGATGTACCTAATCGGGGTGCAACAGCAAGGCTGGT GACATCAGTTGTAATGCAGATATCAATCCACTGAAAATAGGACAAACATCTTCTTCTGTATCTT TCAAAAGTGAAAAATTTCAAGGCACACCAAGAAATGAACTGCAGAACTGCTCTCTGTAGTAATGT TACCTGCTGGTTGAAAGACGTTTACATGAAAGGAGAATACTTTGTTAATGTGACTACCAGAAAT TGGAACGGGACTTTCCATCATCAACGTTCCAGACAGTACAGCTAACGGCAGCTGCAGAAATCA ACACCTATAACCTGAGATATATGTGATTGAAGATAACACTGTTACGATTCCCCTGATGATAAT GAAACCTGATGAGAAAGCCGAAGTACCAACAGGAGTTAATAGGAAGTATAATTGCTGGAATC CTTTGCTGTTAGCTCTGGTTGCAATTTATGGAAGCTCGGCTTCTTCAAAGAAAAATATGAAA AGATGACCAAAAAATCCAGATGAGATTGATGAGACCACAGAGCTCAGTAGCTGAACACAGCAGACC TACCTG	
	ORF Start: ATG at 2	ORF Stop: TGA at 3443

	SEQ ID NO: 42	1147 aa	MW at 125495.9kD
NOV11a, CG165676-01 Protein Sequence	MGPERTGAAPLPILLVLALSQGILNCLAYNVGLPEAKIFSGPSSEQFGYAVQQFINPKGNWLL VGSPPWSGFPENRMGDVYKCPVDLSTATCEKLNLOSTSTIPNVTEMKTNMSLGLILTRNMGTTGGF LTCGPLWAQQCGNQYTTTGVCSDISPDFQLSASFSPATQPCPSLIDVVVVCDESNSIYPWDAVK NFLEKFKVQLDIGPTKTQVGLIQYANNPRVFNLTNYKTKKEEMIVATSQTSQYGGDLTNTFGAI QYARKYAYSAASGRRSATKVMVVVTDGESHDSMLKAVIDQCNDHNLRFGLIAVLGYLNRNAL DTKNLIKELALASIPTERYFFNVSDAALLEKAGLTGEQIFSIIEGTVQGGDNFQMEMSQVGF ADYSSQNDILMLGAVGAFWSGTIVQKTSHGHLIFPKQAFDQILQDRNHSSYLGYSVAAISTGE STHFVAGAPRANYTGQIVLYSVNENGNITVIQHRGDTIGSYFGSVLCSVDVDRDITI TDVLLVG		

APMYMSDLKKKEGRVYLFTIKEGILGQHQLLEGPEGIENTRFGSAIAALSDINMDGFNDVIVGS PLENQNSGAVYIYNHQGTIRTKYSQKILGSDGAFRSHLQYFGRSLDGYGDLNGDSITDVSIGA FGQVVQLWSQSIADVAIEASFTPEKITLVNKNAQIILKLCFSAKFRPTKQNNQVAIVYNITLDA DGFSSRVTSRGLFKENNERCLQKNMVVNQAQSCPEHIYYIQEPSDVVNSLDLRVDISLENPGTS PALEYSETAKVFSIPFHKDCGEDGLCISDLVLDVROIPAAEQPFIVSNQNKRLTFVTLKNK RESAYNTGIVVDFSENLFASFSLPVDGTEVTCQVAASQKSVACDVGYPALKREQQVTFTINFD FNLQNLQNASLSFQALSESQEEKNADNLVNLKIPLLYDAEIHLLTKVTTGSPVPVSMATVIIHIP QYTKEKNPLMYLTGVQTDKAGDISCNADINPLKIGQTSSTSSVSFKSENFRHTKELNCRASCNSV TCWLKDVHMKGEYFVNVTTRIWNGTFASSTFQTVQLTAAAEINTYNPEIYVIEDNTFTIPLMIM KPDEKAEVPTGVIIGSIIAGILLALLALVAILWKLGFPPKRKYKMTKNPDEIDETTELSS
---

Further analysis of the NOV11a protein yielded the following properties shown in Table 11B.

Table 11B. Protein Sequence Properties NOV11a	
SignalP analysis:	Cleavage site between residues 30 and 31
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 5; pos.chg 1; neg.chg 1  H-region: length 30; peak value 9.82  PSG score: 5.42</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): 1.12  possible cleavage site: between 22 and 23</p> <p>&gt;&gt;&gt; Seems to have a cleavable signal peptide (1 to 22)</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 23  Tentative number of TMS(s) for the threshold 0.5:  2  Number of TMS(s) for threshold 0.5: 1  INTEGRAL Likelihood = -13.27 Transmembrane  1100 -1116  PERIPHERAL Likelihood = 0.95 (at 943)  ALOM score: -13.27 (number of TMSs: 1)</p> <p>MTOP: Prediction of membrane topology (Hartmann et al.)  Center position for calculation: 11  Charge difference: -2.0 C(-1.0) - N( 1.0)  N &gt;= C: N-terminal side will be inside</p> <p>&gt;&gt;&gt; membrane topology: type 1a (cytoplasmic tail 1117 to 1147)</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 1 Hyd Moment (75): 10.10  Hyd Moment (95): 5.95 G content: 4  D/E content: 2 S/T content: 2  Score: -6.90</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  R-2 motif at 15 ERT GA</p>

	<p>NUCDISC: discrimination of nuclear localization signals  pat4: none  pat7: none  bipartite: none  content of basic residues: 7.7%  NLS Score: -0.47</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals:  XXRR-like motif in the N-terminus: GPER</p> <p>none</p> <p>SKL: peroxisomal targeting signal in the C-terminus:  none</p> <p>PTS2: 2nd peroxisomal targeting signal: none</p> <p>VAC: possible vacuolar targeting motif: none</p> <p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif:  type 1: none  type 2: none</p> <p>NMYR: N-myristoylation pattern : none</p> <p>Prenylation motif: none</p> <p>memYQRL: transport motif from cell surface to Golgi:  none</p> <p>Tyrosines in the tail:1128</p> <p>Dileucine motif in the tail: none</p> <p>checking 63 PROSITE DNA binding motifs: none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p> <p>checking 33 PROSITE prokaryotic DNA binding motifs:  none</p> <p>NNCN: Reinhardt's method for Cytoplasmic/Nuclear  discrimination  Indication: cytoplasmic  Reliability: 89</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions  total: 0 residues</p> <p>-----</p> <p>Final Results (k = 9/23):</p> <p>55.6 %: endoplasmic reticulum  22.2 %: Golgi  11.1 %: plasma membrane</p>
--	--

	11.1 %: extracellular, including cell wall
	>> indication for CG165676-01 is end (k=9)

A search of the NOV11a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 11C.

5

Table 11C. Geneseq Results for NOV11a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV11a. Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAV07729	Armenian hamster alpha-2 integrin subunit protein - Cricetus migratorius, 1183 aa. [WO9916465-A1, 08-APR-1999]	1..1147 1..1183	1139/1184 (96%) 1141/1184 (96%)	0.0
AAW70542	Integrin alpha-2 chain - Homo sapiens, 1367 aa. [WO9832771-A1, 30-JUL-1998]	1..1098 1..1132	1095/1132 (96%) 1097/1132 (96%)	0.0
ABG29239	Novel human diagnostic protein #29230 - Homo sapiens, 979 aa. [WO200175067-A2, 11-OCT-2001]	206..1147 4..979	937/976 (96%) 940/976 (96%)	0.0
ABB90759	Human Tumour Endothelial Marker polypeptide SEQ ID NO 250 - Homo sapiens, 1179 aa. [WO200210217-A2, 07-FEB-2002]	23..1131 22..1175	466/1182 (39%) 680/1182 (57%)	0.0
ABB90788	Rat Tumour Endothelial Marker polypeptide SEQ ID NO 307 - Rattus sp., 1180 aa. [WO200210217-A2, 07-FEB-2002]	1..1131 1..1176	471/1202 (39%) 678/1202 (56%)	0.0

In a BLAST search of public sequence databases, the NOV11a protein was found to have homology to the proteins shown in the BLASTP data in Table 11D.

10

<b>Table 11D. Public BLASTP Results for NOV11a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV11a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
AAM34795	Integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor) - Homo sapiens (Human), 1181 aa.	1..1147 1..1181	1147/1181 (97%) 1147/1181 (97%)	0.0
P17301	Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GPIa) (Collagen receptor) (VLA-2 alpha chain) (CD49b) - Homo sapiens (Human), 1181 aa.	1..1147 1..1181	1146/1181 (97%) 1147/1181 (97%)	0.0
P53710	Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GPIa) (Collagen receptor) (VLA-2 alpha chain) (CD49b) - Bos taurus (Bovine), 1170 aa (fragment).	12..1147 1..1170	986/1170 (84%) 1069/1170 (91%)	0.0
Q62469	Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GPIa) (Collagen receptor) (VLA-2 alpha chain) (CD49b) - Mus musculus (Mouse), 1178 aa.	1..1147 1..1178	945/1181 (80%) 1040/1181 (88%)	0.0
O42094	ALPHA1 integrin - Gallus gallus (Chicken), 1171 aa.	29..1131 17..1167	456/1179 (38%) 671/1179 (56%)	0.0

PFam analysis indicates that the NOV11a protein contains the domains shown in the Table 11E.

5

<b>Table 11E. Domain Analysis of NOV11a</b>			
<b>Pfam Domain</b>	<b>NOV11a Match Region</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
FG-GAP	45..103	16/65 (25%) 38/65 (58%)	6.1e-05
vwa	174..357	71/208 (34%) 155/208 (75%)	1.8e-63
FG-GAP	434..486	16/64 (25%) 38/64 (59%)	2.2e-06
FG-GAP	488..549	21/66 (32%) 47/66 (71%)	4.7e-13

FG-GAP	551..610	24/67 (36%) 53/67 (79%)	2.2e-17
FG-GAP	615..667	16/66 (24%) 42/66 (64%)	5e-08
integrin_A	1121..1135	7/15 (47%) 14/15 (93%)	0.0055

**Example 12.**

The NOV12 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 12A.

5

Table 12A. NOV12 Sequence Analysis			
	SEQ ID NO: 43	1105 bp	
NOV12a, CG165719-04 DNA Sequence	AGGAGGAAAAACAAGTGTGTGTTGGGGGGAACAGGGGGAAAAGCATTTTTGGTGGATGGTATGA AGCCAGCCATGGAACTGCAGCCGAGGAAAATACTGAACAAAGCCAAGAGAGAAAAGTGAACAG CAGAGCTGAAATGGAAATTGGCAGGTACCACTGGATGTACCCAGGCTCAAAGAACCACAGTAC CATCCCGTGCCAACCTGGGGGACAGGGCTAGCCCTTGAGCAGTCCAGGCTGCTTTGAATGCT GCATCAAGTGTCTGGGAGGAGTCCCTACGCCTCCCTGGTGGCCACCATCCTCTGCTTCTCCGG GGTGGCCTTATTCTGCGGCTGTGGGCATGTGGCTCTCGCAGGCACCGTGGCGATTCTTGAGCAA CACTTCTCCACCAACGCCAGTGACCATGCCCTTGCTGAGCGAGGTGATACAACCTGATGCAGTATG TCATCTATGGAATTGCGTCTTTTCTTCTGTATGGGATCATTCTGTTGGCAGAAGGCTTTTA CACCACAAGTGCAAGTGAAGAACTGCACGGTGAGTTTAAAAACAACCGCTTGTGGCCGATGCATC AGTGGAAATGTTCTGTTTCTCCTCACCTATGTGCTTGGAGTGGCCTGGCTGGGTGTGTTGGTTTCT CAGCGGTGCCCGTGTATGTTCTACAACATATGGTCAACTTGTAAGTCATCAAGTCACCGCA GACCAACGGGACCAACGGTGTGGAGCAGATCTGTGTGGATATCCGACAATACGGTATCATTCT TGGAATGCTTTCCCGGAAAAATATGTGGCTCTGCCCTGGAGAACATCTGCAACACAAACGAGT TCTACATGTCCTATCACCTGTTTCTGTGGCCTGTGCAGGAGCTGGTGGCCACCGTCATTGGCCCT GATCCACTTCTCATGATACTGTCTTCTAAGTGGGCTTACTTAAAGGATGCGAGCAAAATGCAG GCTTACCAGGATATCAAAGCAAAGGAAGAACAGGAACCTGCAAGATATCCAGTCTCGGTCAAAG AACTCAATCTTACACATAAATGTTTGCCAGAGTGTTCGCGCCGACGTATTACAGCTCTG ACAAATCATCAGACAGC		
	ORF Start: ATG at 61		ORF Stop: TAA at 1045

	SEQ ID NO: 44	328 aa	MW at 36219.3kD
NOV12a, CG165719-04 Protein Sequence	MKPAMETAABENTEQSQERKVNRAEMEIGRYHWMYPGSKNHQYHPVPTLGDRASPLSSPGCFE CCIKCLGGVPYASLVATILCFSGVALFCGCHVALAGTVAILEQHFSTNASDHALLSEVIQLMQ YVIYGIASFFFLYGIILLAEGFYTTSVAVKELHGEFKTTACGRCSGMFVFLTYVLGVAWLGVFG FSAVPVFMFYNIWSTCEVIKSPQTNNGTGVQICVDIRQYGLIPWNAFPKGKICGSALENI CNTN EFYMSYHLFIVACAGAGATVIALIHFLMILSSNWAYLKDASKMQAYQDIKAKKEQELQDIQSR KEQLNSYT		

10

	SEQ ID NO: 45	1133 bp	
NOV12b, CG165719-02 DNA Sequence	AGGAGGAAAAACAAGTGTGTGTTGGGGGGAACAGGGGGAAAAGCATTTTTGGTGGATGGTATGA AGCCAGCCATGGAACTGCAGCCGAGGAAAATACTGAACAAAGCCAAGAGAGAAAAGTGAACAG CAGAGCTGAAATGGAAATTGGCAGGTACCACTGGATGTACCCAGGCTCAAAGAACCACAGTAC CATCCCGTGCCAACCTGGGGGACAGGGCTAGCCCTTGAGCAGTCCAGGCTGCTTTGAATGCT GCATCAAGTGTCTGGGAGGAGTCCCTACGCCTCCCTGGTGGCCACCATCCTCTGCTTCTCCGG GGTGGCCTTATTCTGCGGCTGTGGGCATGTGGCTCTCGCAGGCACCGTGGCGATTCTTGAGCAA CACTTCTCCACCAACGCCAGTGACCATGCCCTTGCTGAGCGAGGTGATACAACCTGATGCAGTATG		

	TCATCTATGGAATTGCGTCCTTTTCTCTTGTATGGGATCAATTCGTTGGCAGAAGGCTTTTACACCACAAGTGCAGTGAAAGAACTGCACGGTGAGTTAAAAACAACCGCTTGTGGCCGATGCATCAGTGGAAATGTTCTGTTTTCTCACTATGTGCTTGGAGTGGCCTGGCTGGGTGTGTTTGGTTTCTCAGCGGTGCCCGTGTATGTTCTACAACATATGGTCAACTTGTGAAGTCATCAAGTCACCGCAGACCAACGGGACCACGGTGTGGAGCAGATCTGTGTGGATATCCGACAATACGGTATCATTCCTTGGAAATGCTTTCCCGGAAAAATATGTGGCTCTGCCCTGGAGAACATCTGCAACACAAACGAGTTCTACATGTCTATCACCTGTTTATTGTGGCCTGTGCAGGAGCTGGTGCCACCGCTCATTTGCCCTGCTGATCTACATGATGGCTACTACATATAACTATGCGGTTTTGAAGTTTAAGAGTCGGGAAGATTGCTGCACTAAATTCTAAATTGCATAAGGAGTTTATAGAGCTATGCTCTGTAGCATGAAATATCACTGACACTCCAGACTAAAGCAGAGTCTAGGTTTCTGCAATTTGTACAGTAATTTGTAATAGCTTTGTAACTCACTGCATGTAGATAATAAGATGACTACTGTACA		
	ORF Start: ATG at 61		ORF Stop: TAA at 976

	SEQ ID NO: 46	305 aa	MW at 33537.5kD
NOV12b, CG165719-02 Protein Sequence	MKPAMETAEEENTEQSQERKVNRAEMEIGRYHWMYPGSKNHQYHPVPTLGDRASPLSSPGCFECCIKCLGGVPYASLVATILCFSGVALFCGCGHVALAGTVAILEQHFSTNASDHALLSEVIQLMQYVIYGIASFYFLYGIILLAEFGYTTSAVKELHGEFKTTACGRCSGMFVFLTYVLGVAVLGVFGFSAPVFMFYNWSTCEVIKSPQTNGTTGVEQICVDIRQYGIIPWNAFPKGKICGSALENI CNTNEFYMSYHLFIVACAGAGATVIALLIYMMATTNYAVLKFKSREDCCCTKF		

5

	SEQ ID NO: 47	1182 bp	
NOV12c, CG165719-03 DNA Sequence	AGGAGGAAAAACAAGTGTGTGTTGGGGGGAACAGGGGAAAAGCATTTTTGGTGGATGGTATGAAGCCAGCCATGGAAGCTGCAGCCGAGGAAAATACCTGAACAAAGCCAAGAGAGAAAAGGCTGCTTTGAATGCTGCATCAAGTGTCTGGGAGGAGTCCCTACGCCTCCCTGGTGGCCACCATCCTCTGCTTCTCCGGGTGGCCTTATTCTGCGGCTGTGGGCATGTGGCTCTCGCAGGCACCGTGGCGATTCTTGAGCAACACTTCTCCACCAACGCCAGTGACCATGCCTTGCTGAGCGAGGTGATACAACTGATGCAGTATGTCTATGGAATTGCGTCCTTTTCTTCTTGTATGGGATCAATCTGTTGGCAGAAAGGCTTTTACACCAACAAGTGCAAGTGAAAGAACTGCACGGTGAGTTTAAAAACAACCGCTTGTGGCCGATGCATCAGTGGAAATGTTCTGTTTTCTCACTATGTGCTTGGAGTGGCTGGCTGGGTGTGTTTGGTTTTCTCAGCGGTGCCCGTGTATGTTCTACAAATATGGTCAACTTGTGAAGTCATCAAGTCACCGCAGACCAACGGGACCAACGGGTGTGGAGCAGATCTGTGTGGATATCCGACAATACGGTATCATTCTCTTGGAAATGCTTTCCCGGAAAAATATGTGGCTCTGCCCTGGAGAACATCTGCAACACAAACGAGTTCTACATGTCTATCACTGTTTATTGTGGCCTGTGCAGGAGCTGGTGCCACCGTCAATTGCCCCGATCCACTTCTCATGATAGTCTTCTAACTGGGCTTACTTAAAGGATGCGAGCAAAATGCAGGCTTACCAGGATATCAAGCAAAGGAAGAACAGGAACCTGCAAGATATCCAGTCTCGTCAAAAGAACAACTCAATTCTTACACATAAATGTTTGCCAGAGTGTTCGGCCGACGTATTTACAGCTCTGACAAATCATCAGACAGCTGCTCTGCAGTACAGATGTGTATCCCACCAAACTAATGTAGATGTACAAACACTTCACTGTCTGTCTCAAGCTGCTGGGATGTATCTCTAGGAAAACCTTCCA GTGGGTAAATCTTTTCTTTAGAACAAATATTGGAGTTTATGTTGCCCATTTAAAGGGCACA CTTTACAAATGATCGTCACTTTGGGAT		
	ORF Start: ATG at 61		ORF Stop: TAA at 925

	SEQ ID NO: 48	288 aa	MW at 31670.3kD
NOV12c, CG165719-03 Protein Sequence	MKPAMETAEEENTEQSQERKGFCECCIKCLGGVPYASLVATILCFSGVALFCGCGHVALAGTVAILEQHFSTNASDHALLSEVIQLMQYVIYGIASFYFLYGIILLAEFGYTTSAVKELHGEFKTTACGRCSGMFVFLTYVLGVAVLGVFGFSAPVFMFYNWSTCEVIKSPQTNGTTGVEQICVDIRQYGIIPWNAFPKGKICGSALENI CNTNEFYMSYHLFIVACAGAGATVIALIHFLMILSSNWAYLKDA SKMQAYQDIKAKEEQELQDIQSRSEQLNSYT		

10

	SEQ ID NO: 49	1302 bp	
NOV12d, CG165719-01 DNA Sequence	AGGAGGAAAAACAAGTGTGTGTTGGGGGGAACAGGGGGAAAAGCATTTTTGGTGGATGGTATGA AGCCAGCCATGGAACTGCAGCCGAGGAAAAATACTGAACAAAGCCAAGAGAGAAAAAGTGAACAG CAGAGCTGAAATGGAAATTGGCAGGTACCACTGGATGTACCCAGGCTCAAAGAACCACAGTAC CATCCCGTGCCAACCCCTGGGGGACAGGGCTAGCCCCCTTGAGCAGTCCAGGCTGCTTTGAATGCT GCATCAAGTGTCTGGGAGGAGTCCCCTACGCCTCCCTGGTGGCCACCATCCTCTGCTTCTCCGG GGTGGCCTTATTCTGCGCTGTGGGCATGTGGCTCTCGCAGGCACCGTGGCGATTCTTGAGCAA CACTTCTCCACCAACGCCAGTGACCATGCCTTGCTGAGCGAGGTGATACAACTGATGCAGTATG TCATCTATGGAATTGCGTCTTTTCTTCTGTATGGGATCATTCTGTTGGCAGAAAGCCTTTTA CACCACAAGTGCAGTGAAGAAGTGCACGGTGAGTTTAAAAACAACCGCTTGTGGCCGATGCATC AGTGAATGTTTCTGTTTTCTCACCTATGTGCTTGGAGTGGCCTGGCTGGTGTGTTTGGTTTCT CAGCGGTGCCCCGTGTTTATGTTCTACAACATATGGTCAACTTGTGAAGTCATCAAGTACCCGCA GACCAACGGGACCAACGGGTGTGGAGCAGATCTGTGTGGATATCCGACAATACGGTATCATTCT TGGAATGCTTTCCCGGAAAAATATGTGGCTCTGCCCTGGAGAACATCTGCAACACAAACGAGT TCTACATGTCTTATCACTGTTTCTTGTGGCCTGTGCAGGAGCTGGTGCCACCGTCATTGCCCT GATCCACTTCTCATGATACTGTCTTCTAACTGGGCTTACTTAAAGGATGCGAGCAAAATGCAG GCTTACCAGGATATCAAAGCAAAGGAAGAACAGGAACGCAAGATATCCAGTCTCGGTCAAAG AACAACCTCAATTCTTACACATAAATGTTTGCCAGAGTGTTCGCCCGACGTATTTACAGCTCTG ACAAATCATCAGACAGCTGCTCTGCAGTACAGATGTGTATCCCAACAACTAATGTAGATGTAC AAACACTTCACTGTCTGTCTCAAGCTGCTGGGATGTATCTCTAGAAAACCTTCCAGTGGGTAA ATCTTTTCTTTAGAACAAATATTGGAGGTTTCATGTTGCCCCATTAAAGGGCACACTTTTACA AATGATCGTCATACTTTGGGAT		
	ORF Start: ATG at 61		ORF Stop: TAA at 1045

	SEQ ID NO: 50	328 aa	MW at 36219.3kD
NOV12d, CG165719-01 Protein Sequence	MKPAMETAAEENTEQSQERKVNRAEMEIGRYHWMYPGSKNHQVHPVPTLGDRA SPLSSPGCFE CCIKCLGGVVPYASLVATILCFSGVALFCGCGHVALAGTVAILEQHFSTNASDHALLSEVIQLMQ YVIYGIASFPLYGIILLAEFGYTTSAVKELHGEFKTTACGRCSGMFVFLTYVLGVAVLGVFG FSAVPVFMFYNIWSTCEVIKSPQNTGTTGVEQICVDIRQYGIIPWNAFPGKICGSALENI CNTN EFYMSYHLFIVACAGAGATVIALIHFLMILSSNWAYLKDASKMQAYQDIKAKEQEQLQDIQSRS KEQLNSYT		

5

	SEQ ID NO: 51	929 bp	
NOV12e, CG165719-05 DNA Sequence	AGGAGGAAAAACAAGTGTGTGTTGGGGGGAACAGGGGGAAAAGCATTTTTGGTGGATGGTATGA AGCCAGCCATGGAACTGCAGCCGAGGAAAAATACTGAACAAAGCCAAGAGAGAAAAAGGCTGCTT TGAATGCTGCATCAAGTGTCTGGGAGGAGTCCCCTACGCCTCCCTGGTGGCCACCATCCTCTGC TTCTCCGGGGTGGCCTTATTCTGCGGCTGTGGGCATGTGGCTCTCGCAGGCACCGTGGCGATTCT TTGAGCAACACTTCTCCACCAACGCCAGTGACCATGCCTTGCTGAGCGAGGTGATACAACTGAT GCAGTATGTCTATGGAATTGCGTCTTTTCTTCTGTATGGGATCATTCTGTTGGCAGAA GGCTTTTACACCACAAGTGCAAGTGAAGAAGTGCACGGTGAGTTTAAAAACAACCGCTTGTGGCC GATGCATCAGTGAATGTTGCTTTTCTCACCTATGTGCTTGGAGTGGCCTGGCTGGGTGTGTT TGGTTTCTCAGCGGTGCCGTGTTTATGTTCTACAACATATGGTCAACTTGTGAAGTCATCAAG TCACCGCAGACCAACGGGACCAACGGGTGTGGAGCAGATCTGTGTGGATATCCGACAATACGGTA TCATTCTTGGGAATGCTTTCCCGGAAAAATATGTGGCTCTGCCCTGGAGAACATCTGCAACAC AAACGAGTTCTACATGCTCTATCACCTGTTTCTGTTGGCCTGTGCAGGAGCTGGTGCCACCGTC ATTGCCCTGCTGATCTACATGATGGCTACTACATATAACTATGCCGTTTTGAAGTTTAAGAGTTC GGGAAGATTGCTGCACTAAATTCTAAATTGCATAAGGAGTTTTAGAGAGCTATGCTCTGTAGCA TGAAATATCACTGACACTCCAGAAAGGGCGATT		
	ORF Start: ATG at 61		ORF Stop: TAA at 856

	SEQ ID NO: 52	265 aa	MW at 28988.5kD
NOV12e, CG165719-05 Protein Sequence	MKPAMETAAEENTEQSQERKGCFCCKIKLGGVVPYASLVATILCFSGVALFCGCGHVALAGTVA ILEQHFSTNASDHALLSEVIQLMQYVIYGIASFPLYGIILLAEFGYTTSAVKELHGEFKTTAC GRCSGMFVFLTYVLGVAVLGVFGFSAVPVFMFYNIWSTCEVIKSPQNTGTTGVEQICVDIRQY GIIPWNAFPGKICGSALENI CNTNEFYMSYHLFIVACAGAGATVIALLIYMMATTNYAVLKF SREDCTKF		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 12B.

Table 12B. Comparison of NOV12a against NOV12b through NOV12e.		
Protein Sequence	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV12b	1..298 1..298	285/298 (95%) 291/298 (97%)
NOV12c	1..328 1..288	288/328 (87%) 288/328 (87%)
NOV12d	1..328 1..328	328/328 (100%) 328/328 (100%)
NOV12e	1..298 1..258	245/298 (82%) 251/298 (84%)

5

Further analysis of the NOV12a protein yielded the following properties shown in Table 12C.

Table 12C. Protein Sequence Properties NOV12a	
SignalP analysis:	No Known Signal Sequence Indicated
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 11; pos.chg 1; neg.chg 3  H-region: length 2; peak value 0.00  PSG score: -4.40</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): -10.13  possible cleavage site: between 60 and 61</p> <p>&gt;&gt;&gt; Seems to have no N-terminal signal peptide</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 1  Tentative number of TMS(s) for the threshold 0.5:</p> <p>4  INTEGRAL Likelihood = -6.26 Transmembrane  78 - 94  INTEGRAL Likelihood = -6.74 Transmembrane  130 - 146  INTEGRAL Likelihood = -5.15 Transmembrane  175 - 191  INTEGRAL Likelihood = -8.17 Transmembrane  264 - 280</p>

	<p>PERIPHERAL Likelihood = 3.07 (at 195) ALOM score: -8.17 (number of TMSs: 4)</p> <p>MTOP: Prediction of membrane topology (Hartmann et al.) Center position for calculation: 85 Charge difference: 0.0 C( 0.0) - N( 0.0) N &gt;= C: N-terminal side will be inside</p> <p>&gt;&gt;&gt; membrane topology: type 3a</p> <p>MITDISC: discrimination of mitochondrial targeting seq R content: 0 Hyd Moment(75): 7.96 Hyd Moment(95): 11.32 G content: 0 D/E content: 2 S/T content: 1 Score: -5.88</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq cleavage site motif not found</p> <p>NUCDISC: discrimination of nuclear localization signals pat4: none pat7: none bipartite: none content of basic residues: 6.1% NLS Score: -0.47</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals: none</p> <p>SKL: peroxisomal targeting signal in the C-terminus: none</p> <p>PTS2: 2nd peroxisomal targeting signal: none</p> <p>VAC: possible vacuolar targeting motif: none</p> <p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif: type 1: none type 2: none</p> <p>NMYR: N-myristoylation pattern : none</p> <p>Prenylation motif: none</p> <p>memYQRL: transport motif from cell surface to Golgi: none</p> <p>Tyrosines in the tail: none</p> <p>Dileucine motif in the tail: none</p> <p>checking 63 PROSITE DNA binding motifs: none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p>
--	--

none

NNCN: Reinhardt's method for Cytoplasmic/Nuclear  
discrimination

Indication: cytoplasmic

Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions

290 A 0.60

291 Y 0.75

292 L 0.75

293 K 0.88

294 D 0.93

295 A 0.98

296 S 0.99

297 K 1.00

298 M 1.00

299 Q 1.00

300 A 1.00

301 Y 1.00

302 Q 1.00

303 D 1.00

304 I 1.00

305 K 1.00

306 A 1.00

307 K 1.00

308 E 1.00

309 E 1.00

310 Q 1.00

311 E 1.00

312 L 1.00

313 Q 1.00

314 D 1.00

315 I 1.00

316 Q 1.00

317 S 1.00

318 R 1.00

319 S 1.00

320 K 1.00

321 E 1.00

322 Q 1.00

323 L 1.00

324 N 1.00

325 S 0.97

326 Y 0.96

327 T 0.87

total: 38 residues

-----  
Final Results (k = 9/23):

55.6 %: endoplasmic reticulum

44.4 %: mitochondrial

>> indication for CG165719-04 is end (k=9)

A search of the NOV12a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 12D.

<b>Table 12D. Geneseq Results for NOV12a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV12a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
ABG70364	Novel human thrombopoietin variant protein, NV-23 - Homo sapiens, 279 aa. [US2002068342-A1, 06-JUN-2002]	27..298 1..272	259/272 (95%) 265/272 (97%)	e-155
AAY09510	Human M6b1 protein - Homo sapiens, 265 aa. [WO9921982-A1, 06-MAY-1999]	1..298 1..258	245/298 (82%) 251/298 (84%)	e-138
AAW39215	Human M6 protein - Homo sapiens, 278 aa. [JP10014577-A, 20-JAN-1998]	61..328 13..278	155/276 (56%) 209/276 (75%)	2e-86
ABG02005	Novel human diagnostic protein #1996 - Homo sapiens, 541 aa. [WO200175067-A2, 11-OCT-2001]	49..294 289..533	134/246 (54%) 173/246 (69%)	1e-76
AAR95171	Murine CNS myelin membrane proteolipid protein isoform DM20 - Mus musculus, 242 aa. [EP685558-A1, 06-DEC-1995]	61..294 2..234	130/234 (55%) 169/234 (71%)	8e-76

5

In a BLAST search of public sequence databases, the NOV12a protein was found to have homology to the proteins shown in the BLASTP data in Table 12E.

<b>Table 12E. Public BLASTP Results for NOV12a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV12a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q8N956	Hypothetical protein FLJ38338 - Homo sapiens (Human), 328 aa.	1..328 1..328	328/328 (100%) 328/328 (100%)	0.0
Q9JI65	Neuronal membrane glycoprotein M6-B - Mus musculus (Mouse), 328 aa.	1..328 1..328	321/328 (97%) 324/328 (97%)	0.0

P35803	Neuronal membrane glycoprotein M6-b (M6b) - Mus musculus (Mouse), 288 aa.	1..328 1..288	284/328 (86%) 287/328 (86%)	e-162
Q98ST3	Myelin PLP-related membrane protein DM gamma1 - Xenopus laevis (African clawed frog), 269 aa.	61..328 2..269	237/268 (88%) 250/268 (92%)	e-141
Q8UUS8	DMgamma2 - Brachydanio rerio (Zebrafish) (Danio rerio), 268 aa.	61..328 2..268	218/268 (81%) 244/268 (90%)	e-132

PFam analysis indicates that the NOV12a protein contains the domains shown in the Table 12F.

Table 12F. Domain Analysis of NOV12a			
Pfam Domain	NOV12a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Myelin_PLP	61..305	175/288 (61%) 243/288 (84%)	2.3e-196

5

### Example 13.

The NOV13 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 13A.

Table 13A. NOV13 Sequence Analysis			
	SEQ ID NO: 53	1201 bp	
NOV13a, CG167488-02 DNA Sequence	TTTGATCTGAAGACTAGGGGACAATGGATATCATAGAGACAGCAAACTTGAAGAACATTGGGA AAATCAACCCAGTGATCTACGAACACTTATGCAAGACCCGCTGAACCTGTTGAAGAAGAAAAC AAAAATGGCAATGGTAAACCCAGAGCTTATCCAGTGGGCTGCGAAAAGGCACCAAAAAGTACC CGGACTATATCCAAATTGCTATGCCACTGAATCAAGGAACAAATTTCCACTAGAGTGGTGGAA AACGGGCATTGCCCTTCATATATGCAGTTTTCAACCTCGTCTTGACAACCGCTCATGATCAGATT GTACATGAGAGGGTCCCTCCCAAGGAGCTTAGCCCTCCACTCCAGACAAGTTTTTGGATTACA TTGATAGGGTGAAATGGGCATTTTCTGTATCAGAAATAAATGGGATTATATTAGTTGGATTATG GATCACCAGTGGCTGTTTCTGAGATACAAGTCAATAGTGGGACGCAGATTCTGTTTTATATT GGAACTTTATACCTGTATCGCTGCATTACAATGTATGTTACTACTCTACCTGTGCCTGGAATGC ATTCCAGTGTGCTCCAAAGCTCAATGGAGACTCTCAGGCAAAAGTCAACGGATTCTACGATT GATTCTGGTGGTGGATTGTCCATAACTGGATCACAATCTTATGTGGAGACTTCTCTTCAGC GGTCACACGGTTACGCTGACACTGACTTATTTGTTTCATCAAAGAAATATTGCGCTCGTCACTTCT GGTGGTATCATTTAATCTGCTGGCTGCTGAGTGTGCTGCCGGATCATCTGCATTCTTGTAGCACA CGAACACTACACTATCGATGTGATCATTGCTTATATATCACAACAGCACTGTTTGGTGGTAC CATTCAATGGCCAATGAAAAGAACTTGAAGGTCTCTTCACAGACTAATTTCTTATCTCGAGCAT GGTGGTTCCCATCTTTTATTTTGTGAGAAAATGTACAAGGCTCAATTCCTGCTGCTTCTC CTGGCCGCTGTCTTGGCTCCTGGCTGCTTCAAATCATCATGCAAAAAGTATTACGGGTTTCAG AAGATTGGTGAAGACAATGAGAAATCGACCTGAGGAGCAAAACAAAGGCATCAGCTCTTACACC AAAAGAGTTAAGCTGTAAACCAAGAAGGGCGATTCCAGCACACTGCGC		
	ORF Start: ATG at 24		ORF Stop: TGA at 1119

	SEQ ID NO: 54	365 aa	MW at 42279.7kD
NOV13a, CG167488-02 Protein Sequence	MDI IETAKLEEHLNQPSDPTNTYARPAEPVEEENKNGNGKPKSLSSGLRKGTKKYPDYIQIAM PTESRNKFPLEWWTGIAFIYAVFNLVLTVMITVTVHERVPPKELSPPLPDKFFDYIDRVKWAF SVSEINGIILVGLWITQWLFLRYKSVIGRRFCFIIGTLYLRCITMYVTTLVPVGMHFFQCAPKL NGDSQAKVQRIILRLISGGGLSITGSHILCGDFLPSGHTVTLTLTYLFIKEYSPRHFWWYHLICW LLSAAGIICILVAHEHYTIDVIIAYYITTRLFWWYHSMANEKNLKVSSQTNFLSRAWWFPFIYF FEKNVQGSIPCCFSWPLSWPPGCFKSSCKKYSRVQKIGEDNEKST		

	SEQ ID NO: 55	1893 bp	
NOV13b, CG167488-01 DNA Sequence	CGGAGCTACCTTATAAAGACCATCTGTACATCCACTGTGAAATGGAGTTTCAAATCACAAGCT TCTTTCCCACTGAACATAAGACTAGGAGCACATATGGAAGAGTAAAGTTGAAGGGAATTTGGA TGATGATTGGCAAGATGCTGTGGGATAGTAACATCTTTTGGAGGGAAGAATTGGCTTCCTTC TTGAAAGTGGTGAAGGTACAGCATATAGCTGCATGGAAGAAACAGTAATCGGATGGCTACCTTC TACATTTTGTATTAGGAAACAAAGTCCATTGTAAGAGTCCATGTTGATCTTGGAAATAGAAGGA TTGAAAAAAGCTAAATTTCCCAAAGAACAAAGAACTTGACCATCTCCTTTTGTATCTGAAGACT AGGGGACAATGGATATCATAGAGACAGCAAACTTGAAGAACATTTGGAAAAATCAACCCAGTGA TCCTACGAAACATTATGCAAGACCCGCTGAACCTGTTGAAGAAAGAAAAAATAATGGCAATTGG TAAACCCAAGAGCTTATCCAGTGGGCTGCGAAAAGGCACCAAAAAGTACCGGACTATATCCAA ATTGCTATGCCCCTGAATCAAGGAACAAATTTCCACTAGAGTGGTGGAAAAACGGGCATTGCCT TCATATATGCAGTTTCAACCTCGTCTTGACAACCGCTCATGATCACAGTTGTACATGAGAGGGT CCCTCCCAAGGAGCTTAGCCCTCCACTCCAGACAAGTTTTTGTATTACATTGATAGGGTGAAA TGGGCATTTCTGTATCAGAAATAAATGGGATTATATTAGTTGGATTATGGATCACCCAGTGGC TGTTTCTGAGATACAAGTCAATAGTGGGACGCAGATTCTGTTTTATTATTGGAACCTTTATACCT GTATCGCTGCATTACAATGTATGTACTACTCTACCTGTGCTGGAATGCATTTCCAGTGTGCT CCAAAGCTCAATGGAGACTCTCAGGCAAAAGTTCAACGGATTCTACGATTGATTTCTGGTGGTG GATTGTCCATAACTGGATCACATATCTTATGTGGAGACTTCCCTCTCAGCGGATCACACGGTTAC GCTGACACTGACTTATTTGTTTATCAAGAAGATTGCGCTCGTCACTTCTGGTGGTATCATTTA ATCTGCTGGCTGCTGAGTGTGCGGGGATCATCTGCATTCTTGTAGCACACGAACACTACACTA TCGATGTGATCATTTGCTTATTATATCAACAACGACTGTTTTGGTGGTACCATTCAATGGCCAA TGAAAAGAACTTGAAGGTCTCTTCAAGACTAATTTCTTATCTCGAGCATGGTGGTTCCCATC TTTTATTTTTTTGAGAAAAATGTACAAGGCTCAATTCCTTGCTGCTTCTCCTGGCCGCTGTCTT GGCCTCCTGGCTGCTTCAAATCATCATGCAAAAAGTATTACGGGTTTCAAGATTGGTGAAGA CAATGAGAAATCGACCTGAGGAGCAAAACAAAGGCAACAGCTCTTACACCAAAGAGTTAACGC TGTAACCAAAGGTATAGTTTTGTTTTTTATTTTAGGAGAACTGACTGGTAAATGAAGAAATGGA GCAAAATTTGTGTAAACGATTAGAAAGATGAACAAAGTATTGCCCTTTGACTGGTTTCTTCTT CATCCTGAGAAAGATACATTCTCTGAGCTCTTTCATTTCATTGGTGACAGCCCCACCCCGGG ACTTTACTAATGAGCTTGTAAAGAGGTGCCAAAGACATATTCCTCCTTTCTTTATTCTTTCT CCACCAAACCTCTACTTCAGAAATTTTTTCAGGATATTTTCAGCCCAAGGTCAGAAGAATGT GTTAATATTTTAAATAAAATATCTGGACATCTACAAA		
	ORF Start: ATG at 463		ORF Stop: TGA at 1489

5

	SEQ ID NO: 56	342 aa	MW at 39679.3kD
NOV13b, CG167488-01 Protein Sequence	MQDPLNLLKKTKMAIGKPKSLSSGLRKGTKKYPDYIQIAMPTESRNKFPLEWWTGIAFIYAV FNLVLTVMITVTVHERVPPKELSPPLPDKFFDYIDRVKWAFSVSEINGIILVGLWITQWLFLRY KSVIGRRFCFIIGTLYLRCITMYVTTLVPVGMHFFQCAPKLNGDSQAKVQRIILRLISGGGLSIT GSHILCGDFLPSGHTVTLTLTYLFIKEDSPRHFWWYHLICWLLSAAGIICILVAHEHYTIDVII AYYITTRLFWWYHSMANEKNLKVSSQTNFLSRAWWFPFIYFFFEKNVQGSIPCCFSWPLSWPPGC FKSSCKKYSRVQKIGEDNEKST		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 13B.

10

Table 13B. Comparison of NOV13a against NOV13b.

Protein Sequence	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV13b	27..365 4..342	327/339 (96%) 331/339 (97%)

Further analysis of the NOV13a protein yielded the following properties shown in Table 13C.

Table 13C. Protein Sequence Properties NOV13a	
SignalP analysis:	No Known Signal Sequence Indicated
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 11; pos.chg 1; neg.chg 4  H-region: length 2; peak value 0.00  PSG score: -4.40</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): -13.01  possible cleavage site: between 25 and 26</p> <p>&gt;&gt;&gt; Seems to have no N-terminal signal peptide</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 1  Tentative number of TMS(s) for the threshold 0.5:  3  INTEGRAL Likelihood = -6.21 Transmembrane  81 - 97  INTEGRAL Likelihood = -1.59 Transmembrane  133 - 149  INTEGRAL Likelihood = -9.98 Transmembrane  253 - 269  PERIPHERAL Likelihood = 1.11 (at 160)  ALOM score: -9.98 (number of TMSs: 3)</p> <p>MTOP: Prediction of membrane topology (Hartmann et al.)  Center position for calculation: 88  Charge difference: -1.5 C(-0.5) - N( 1.0)  N &gt;= C: N-terminal side will be inside</p> <p>&gt;&gt;&gt; membrane topology: type 3a</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 0 Hyd Moment (75): 4.36  Hyd Moment (95): 8.65 G content: 0  D/E content: 2 S/T content: 0  Score: -6.99</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  cleavage site motif not found</p> <p>NUCDISC: discrimination of nuclear localization signals</p>

	pat4: none pat7: none bipartite: none content of basic residues: 10.7% NLS Score: -0.47  KDEL: ER retention motif in the C-terminus: none  ER Membrane Retention Signals: KQXX-like motif in the C-terminus: NEKS  SKL: peroxisomal targeting signal in the C-terminus: none  PTS2: 2nd peroxisomal targeting signal: none  VAC: possible vacuolar targeting motif: none  RNA-binding motif: none  Actinin-type actin-binding motif: type 1: none type 2: none  NMYR: N-myristoylation pattern : none  Prenylation motif: none  memYQRL: transport motif from cell surface to Golgi: none  Tyrosines in the tail: none  Dileucine motif in the tail: none  checking 63 PROSITE DNA binding motifs: none  checking 71 PROSITE ribosomal protein motifs: none  checking 33 PROSITE prokaryotic DNA binding motifs: none  NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination Indication: cytoplasmic Reliability: 94.1  COIL: Lupas's algorithm to detect coiled-coil regions total: 0 residues  ----- Final Results (k = 9/23):  55.6 %: endoplasmic reticulum 33.3 %: mitochondrial 11.1 %: vesicles of secretory system  >> indication for CG167488-02 is end (k=9)
--	---

A search of the NOV13a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 13D.

Table 13D. Geneseq Results for NOV13a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB09578	Human cytochrome constitutive protein 45 - Homo sapiens, 413 aa. [CN1333280-A, 30-JAN-2002]	1..354 61..409	215/354 (60%) 272/354 (76%)	e-130
AAM41726	Human polypeptide SEQ ID NO 6657 - Homo sapiens, 430 aa. [WO200153312-A1, 26-JUL-2001]	1..354 78..426	215/354 (60%) 272/354 (76%)	e-130
AAM39940	Human polypeptide SEQ ID NO 3085 - Homo sapiens, 413 aa. [WO200153312-A1, 26-JUL-2001]	1..354 61..409	215/354 (60%) 272/354 (76%)	e-130
AAG81320	Human AFP protein sequence SEQ ID NO:158 - Homo sapiens, 222 aa. [WO200129221-A2, 26-APR-2001]	136..354 1..218	150/219 (68%) 181/219 (82%)	2e-93
ABB60637	Drosophila melanogaster polypeptide SEQ ID NO 8703 - Drosophila melanogaster, 384 aa. [WO200171042-A2, 27-SEP-2001]	97..330 1..240	107/244 (43%) 155/244 (62%)	5e-57

5

In a BLAST search of public sequence databases, the NOV13a protein was found to have homology to the proteins shown in the BLASTP data in Table 13E.

Table 13E. Public BLASTP Results for NOV13a				
Protein Accession Number	Protein/Organism/Length	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q8NHU3	Similar to putative - Homo sapiens (Human), 365 aa.	1..365 1..365	365/365 (100%) 365/365 (100%)	0.0
Q9D4B1	4933405A16Rik protein - Mus musculus (Mouse), 270 aa.	96..365 1..270	260/270 (96%) 267/270 (98%)	e-162
Q8VCQ6	Hypothetical 48.7 kDa protein - Mus musculus (Mouse), 413 aa.	1..354 61..409	216/354 (61%) 274/354 (77%)	e-131

CAC38570	Sequence 157 from Patent WO0129221 - Homo sapiens (Human), 222 aa.	136..354 1..218	150/219 (68%) 181/219 (82%)	6e-93
Q9DA37	1700010P07Rik protein - Mus musculus (Mouse), 478 aa.	68..340 204..470	123/273 (45%) 178/273 (65%)	2e-69

**Example 14.**

The NOV14 clone was analyzed, and the nucleotide and encoded polypeptide  
5 sequences are shown in Table 14A.

Table 14A. NOV14 Sequence Analysis			
	SEQ ID NO: 57	1785 bp	
NOV14a, CG173318-01 DNA Sequence	GTCGCCAGCTGAGCGGTTTGTAAGTTTGGGTCGCAGTATGCTAGAATTTGAGGCTCCCTTC TGATGAAAATTGAGCTGTCCATGCAGCCATGGAACCGGGTTACAGCAGTGAGGGGGCCACGGC TCAAGAACTTACACATGTCCAAAATGATTGAGATGGAGCAGGCGGAGGGCCAGCTGTCTGAG TTAGACCTGCTAGCCAGTATGTTCCCTGGTGAGAATGAGCTCATAGTGAATGACCAGCTGGCTG TAGCAGAACTGAAAGATTGTATTGAAAAGAGACAATGGAGGGGCGATCTTCAAAGTCTACTT TACTATCAATATGAACCTGGATGTATCTGACGAAAAAATGGTAATTCAGTTTGTCTTTAGAGG GATTGAAACATGTTGAGACTTAAACATTGGTTAGTGCACTTTTCTTCTCTCTTTAATCAGG CGATGTTTCTCTGGCCTGTATTCTTCCCTTTAAATACCCGGCAGTTCTGCCTGAAATTAAGT CAGATCAGTATTATTGAGTAGATCCAGCAGACTCAGCTGAACACAGATCTGACTGCATTCCTG CAAAAACATTGTCATGGAGATGTTGTATACTGAATGCCACAGAGTGGGTTAGAGAACACGCCT CTGGCTATGTCAGCAGAGATACTTCATCTTCAACCACAGGAAGCACAGTCCAGTCAGTTGA CCTCATCTTCACGAGACTCTGGATCTACAGCCATCATATCTATAACAAATGCAAAAGAAAGAA ATTCTAGAGTGGGCAAAGGAGCTTTCCTGTCTGGGTTAGCATGCCTGGAAAAACCTGGTGTG TTTGTGTGGAAGGCCCAAAAGTGCTGTGAAGAATTCTGGTCAAGACTCAGAAAAATTAAGT GAAGAGAATTTAATTCGCCATCGAGAAGACATTCCTTTGATGGTACAAATGATGAACCGGAA AGACAAAGGAAATTTTCCATTTTGAAGAAAAAGTGTTCAGTGTAAATGGAGCCAGGGGAAAC ACATGGACTTTGGTCACTCTATCAGTTCTTAAACACCAAGGATGTGGGGATGTTTCCAGAT GTTCTTTGGGTGAGAAGGACAATGACATCAAGAGTAGTTGAAAGTATCTTGCCACTGTTGGCCT TTTGATTTTTTTTCCACTTTTCTTGAAGATTAAAGTAATTTATTTAGTTCCATTCTAGA ATGTTGGGGAGTGGGGCAAGAAAAAATAGTATAGCTGAAATGCATCTGTTAAAAATGTCTATG ATTGAAAGCAGAACTGAGTTCAAATTACAACCTTAAATTTGTTGTTAGATATTTCTTCACATA TCAGCTGCCCATTTTGAAGAAAGAAATATCCATAAAGGTAATGTTGGTGCTCCAATTGGCCAGC CATTCCCAACCCCTTCTCCCTTACCTGCCTTCACTAAAGAACCCAGAAAAGCTAATTGCTCCC CTTTCAGCCTCTGTTGCAACTAACAACCTCTCAGTGGCCTCAGGACACAGCTTTGGCCTTGGGAA TTCTGGGAAACTTTTACTTCTCTGATTAAAGATACATATGCAGTAGGCCACCTCTCCCCCCC TTACTGCCATAAACACCAAGTGATGACTGGAGCTGGAGGAGTTATTTGAACACGACGGAAGG GCCAAGAGAACCAAGAGATGCCAGTTGCCACATTGTTGAGCTGCTGACCCACACAGCCATT GCCTGTCTCTAAACATCTTATGAAATAAAACCAATTTGTTTAAAAAATAAAAAA		
	ORF Start: ATG at 394		ORF Stop: TGA at 1111

	SEQ ID NO: 58	239 aa	MW at 27409.3kD
NOV14a, CG173318-01 Protein Sequence	MLRLKTLVSALFLLLFNQMFSILCLPKYPAVLPEITVRSVLLSRSSQQTQLNLDLTAFLQKH CHGDVCLLNATEWVREHASGYVSRDTSSTPTGSTVQSVDLIFTRLWIYSHHIYKCKRKNILE WAKELSLSGFSMPGKPGVVCVEGPGQSACEFWSRLRLNWKRLILRHREDIPFDGTNDETERQR KFSIFEKVFVSVNGARGNHMDFGQLYQFLNTKGCQDVFQMFVGVGQ		

Further analysis of the NOV14a protein yielded the following properties shown in Table 14B.

Table 14B. Protein Sequence Properties NOV14a	
SignalP analysis:	Cleavage site between residues 23 and 24
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 5; pos.chg 2; neg.chg 0  H-region: length 24; peak value 10.88  PSG score: 6.47</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): -2.51  possible cleavage site: between 24 and 25</p> <p>&gt;&gt;&gt; Seems to have no N-terminal signal peptide</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 1  Tentative number of TMS(s) for the threshold 0.5:  1  Number of TMS(s) for threshold 0.5: 1  INTEGRAL Likelihood = -4.78 Transmembrane  11 - 27  PERIPHERAL Likelihood = 5.41 (at 133)  ALOM score: -4.78 (number of TMSs: 1)</p> <p>MTOP: Prediction of membrane topology (Hartmann et al.)  Center position for calculation: 18  Charge difference: -2.0 C( 1.0) - N( 3.0)  N &gt;= C: N-terminal side will be inside</p> <p>&gt;&gt;&gt; membrane topology: type 2 (cytoplasmic tail 1 to 11)</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 1 Hyd Moment(75): 4.83  Hyd Moment(95): 3.70 G content: 0  D/E content: 1 S/T content: 3  Score: -3.64</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  R-2 motif at 57 SRS QQ</p> <p>NUCDISC: discrimination of nuclear localization signals  pat4: none  pat7: none  bipartite: none  content of basic residues: 11.7%  NLS Score: -0.47</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals:  XXRR-like motif in the N-terminus: LRLK</p>

	<p>none</p> <p>SKL: peroxisomal targeting signal in the C-terminus: none</p> <p>PTS2: 2nd peroxisomal targeting signal: none</p> <p>VAC: possible vacuolar targeting motif: none</p> <p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif: type 1: none type 2: none</p> <p>NMYR: N-myristoylation pattern : none</p> <p>Prenylation motif: none</p> <p>memYQRL: transport motif from cell surface to Golgi: none</p> <p>Tyrosines in the tail: none</p> <p>Dileucine motif in the tail: none</p> <p>checking 63 PROSITE DNA binding motifs: none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p> <p>checking 33 PROSITE prokaryotic DNA binding motifs: none</p> <p>NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination indication: cytoplasmic Reliability: 94.1</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions total: 0 residues</p> <p>-----</p> <p>Final Results (k = 9/23):</p> <p>34.8 %: mitochondrial 30.4 %: cytoplasmic 13.0 %: Golgi 8.7 %: endoplasmic reticulum 4.3 %: vacuolar 4.3 %: extracellular, including cell wall 4.3 %: vesicles of secretory system</p> <p>&gt;&gt; indication for CG173318-01 is mit (k=23)</p>
--	--

A search of the NOV14a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 14C.

<b>Table 14C. Geneseq Results for NOV14a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV14a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAE15253	Human RNA metabolism protein-16 (RMEP-16) - Homo sapiens, 319 aa. [WO200183524-A2, 08-NOV-2001]	19..239 99..319	221/221 (100%) 221/221 (100%)	e-131
AAM78405	Human protein SEQ ID NO 1067 - Homo sapiens, 319 aa. [WO200157190-A2, 09-AUG-2001]	19..239 99..319	221/221 (100%) 221/221 (100%)	e-131
AAM79389	Human protein SEQ ID NO 3035 - Homo sapiens, 354 aa. [WO200157190-A2, 09-AUG-2001]	19..236 137..354	215/218 (98%) 216/218 (98%)	e-127
ABB11888	Human novel protein, SEQ ID NO:2258 - Homo sapiens, 354 aa. [WO200157188-A2, 09-AUG-2001]	19..236 137..354	215/218 (98%) 216/218 (98%)	e-127
AAB58229	Lung cancer associated polypeptide sequence SEQ ID 567 - Homo sapiens, 305 aa. [WO200055180-A2, 21-SEP-2000]	19..167 103..251	147/149 (98%) 147/149 (98%)	9e-84

5

In a BLAST search of public sequence databases, the NOV14a protein was found to have homology to the proteins shown in the BLASTP data in Table 14D.

<b>Table 14D. Public BLASTP Results for NOV14a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV14a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
P57060	Protein C21orf6 (GL011) - Homo sapiens (Human), 319 aa.	19..239 99..319	221/221 (100%) 221/221 (100%)	e-130
Q9DCJ3	Open reading frame 5 - Mus musculus (Mouse), 244 aa.	21..239 26..244	182/219 (83%) 192/219 (87%)	e-105
Q99M03	Similar to open reading frame 5 - Mus musculus (Mouse), 290 aa.	21..239 72..290	182/219 (83%) 192/219 (87%)	e-105

Q9JLH4	Orf5 protein - Mus musculus (Mouse), 291 aa.	21..239 73..291	181/219 (82%) 192/219 (87%)	e-105
Q9D9S3	1700030C20Rik protein - Mus musculus (Mouse), 292 aa.	23..239 72..288	85/222 (38%) 127/222 (56%)	4e-38

**Example 15.**

The NOV15 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 15A.

Table 15A. NOV15 Sequence Analysis			
	SEQ ID NO: 59	1776 bp	
NOV15a, CG50970-06 DNA Sequence	CACCGGATCCACCATGTCCGCGCTGCGACCTCTCCTGCTTCTGCTGCTGCCTCTGTGTCCCGGT CCTGGTCCCGGACCCGGGAGCGAGGCAAGGTCACCCGGAGTTGTGCAGAGACCCGGCAGGTGC TGGGGGCCCGGGGATATAGCTTAAACCTAATCCCTCCCGCCCTGATCTCAGGTGAGCACCTCCG GGTCTGTCCCCAGGAGTACACCTGCTGTTCAGTGAGACAGAGCAGAGGCTGATCAGGGAGACT GAGGCCACCTTCCGAGGCCTGGTGGAGGACAGCGGCTCCTTCTGGTTCACACACTGGCTGCCA GGCACAGAAAATTTGATGAGTTTTTCTGGAGATGCTCTCAGTAGCCAGCACTCTCTGACCCA GCTCTTCTCCCACTCCTACGGCCGCTGTATGCCAGCACGCCCTCATATTCAATGGCCTGTTCT TCTCGGCTGCGAGACTTCTATGGGGAATCTGGTGGGGGTTGGATGACACCCCTGGCGGATTCT GGGCACAGCTCCTGGAGAGAGTGTTCCTGCTGCTGCACCCACAGTACAGCTTCCCCCTGACTA CCTGCTCTGCCTCTCACGCTTGGCCTCATCTACCGATGGCTCTCTGCAGCCCTTTGGGGACTCA CCCCGCCGCTCCGCTGCGATAACCCGGACCCCTGGTGGCTGCCCGAGCCCTTTGTGCGAGGGCC TGGAGACTGGAAGAAATGTGGTCAGCGAAGCGCTTAAGGTGCCGGTGTCTGAAGGCTGCAGCCA GGCTCTGATGCGTCTCATCGGCTGTCCCTGTGCGGGGGGTCCCTCTGATTGCCCTGCCAG GGCTTCTGCCTCAACGTGGTTCGTGGCTGTCTCAGCAGCAGGGGACTGGAGCCTGACTGGGGCA ACTATCTGGATGGTCTCCTGATCCTGGCTGATAAGCTCCAGGGCCCCCTTTCTCTTGGAGCTGAC GGCCGAGTCCATTGGGGTGAAGATCTCGAGGGTTTGATGTACCTGCAGGAAAACAGTGCAGAA GTGTCCGCCAGGTGTTTCAGGAGTGCGGCCCCCCCCGACCCGGTGCTGCCCGCAACCGTCGAG CCCCGCCGCCCGGGAAGAGGCGGGCCGCTGTGGTGCATGGTGACCGAGGAGGAGCGGCCAC GACGGCCGCGAGGACCAACCTGCACCGGCTGGTGTGGGAGCTCCGCGAGCGTCTGGCCCGGATG CGGGGCTTCTGGGCCCGGCTGTCCCTGACGGTGTGCGGAGACTCTCGCATGGCAGCGGACGCCT CGCTGGAGGCGGCGCCTGCTGGACCGGAGCCGGGCGGGCGGGTACTTGCCGCCAGTGGTCCG GGGCTCCCCGGCCGAGCAGGTCAACAACCCGAGCTCAAGGTGGACGCCTCGGGCCCCGATGTC CCGACACGGCGGCTCGACTACAGCTCCGGGCGGCCACGGCCAGAATGAAAACGGCCGCACTGG GACACGACCTGGACGGGACGAGGACGCGGATGAGGATGCCAGCGGCTCTGGAGGGGGACAGCAGTA TGCAGATGACTGGATGGCTGGGCTGTGGCTCCCCAGCCCGGCTCCTCGGCTCCATACCTT CCTAGAAGGGATGGTTCTGGGGGCAAGGAGGAGTGGCAGTGCCCGCTACAACAGGGGCCGA GCAGGAGTGGGGGGCATCTATTGGTTTTTCACACCAAAACCATCCTATTCTCTCCCTCTCAGC CCTGGCCCTGCTTGGACCTCGACTCGAGGGCAAGGGCGAATTCAGCA		
	ORF Start: ATG at 14		ORF Stop: at 1751

	SEQ ID NO: 60	579 aa	MW at 62828.7kD
NOV15a, CG50970-06 Protein Sequence	MSALRPLLLLLLPLCPGPGPGSEAKVTRSCAETQVLGARGYSNLNIPPALISGEHLRVCPQ EYTCSSSETEQRLIRETEATFRGLVEDSGSFLVHTLAARHRKFDEFLEMLSVAQHSLTQLFSH SYGRLYAQHALIFNGLFSRLRDFYGESGELDDTLADFWAQLLERVFPLLHPQYSFPPDYLLCL SRLASSTDGSLQPFPGDSPRRLRLQITRTLVAARAFVQGLETRNVVSEALKVPVSEGCQSALMR LIGCPLCRGVPSLMPQCQGFCLNVVRGCLSSRGLEPDWGNLYDLGLLILADKLQGPFSFELTAESI GVKISEGLMYLQENSARKVSAQVFQECGPPDPVPARNRAPPREEAGRLWSMVTEERPTTAAG TNLHRLVWELRERLARMRGFWARLSLTVCDSRMAADASLEAAPCWTGAGRGRYLPVVGGS PAEQVNNPELVKPDASGPDVPTRRRLQLRAATARMKTAALGHDLDGQDADEADASGGGQYADDW MAGAVAPPARPPPPYPPRRDGSCKGGGGSARYNQGRSRSGGASIGFHTQTILLISLSALAIL		

	GPR
--	-----

	SEQ ID NO: 61	1785 bp	
NOV15b, CG50970-01 DNA Sequence	ATGTCCGCGCTGCGACCTCTCCTGCTTCTGCTGCTGCCTCTGTGTCCCGGTCTGGTCCCGGAC CCGGGAGCGAGGCAAGGTCAACCGGAGTTGTGCAGAGACCCGCGAGGTGCTGGGGGCCCGGGG ATATAGCTTAAACCTAATCCCTCCCGCCCTGATCTCAGGTGAGCACCTCCGGGTCTGTCCCCAG GAGTACACCTGCTGTTCCAGTGAGACAGAGCAGAGGCTGATCAGGGAGACTGAGGCCACCTTCC GAGGCTGGTGGAGGACAGCGGCTCCTTTCTGGTTCACACACTGGCTGCCAGGCACAGAAAATT TGATGAGTTTTTCTGGAGATGCTCTCAGTAGCCCACTCTCTGACCCAGCTCTTCTCCAC TCCTACGGCCGCTGTATGCCAGCAGCCCTCATATTCAATGGCTGTCTCTCGGCTGCGAG ACTTCTATGGGAATCTGGTGAGGGTTGGATGACACCTGGCGGATTTCTGGGCACAGCTCCT GGAGAGAGTGTCCCGCTGCTGCACCCACAGTACAGCTTCCCCCTGACTACCTGCTCTGCCTC TCACGCTTGGCTCATCTACCGATGGCTCTCTGCAGCCCTTTGGGGACTCACCCCGCCGCTCC GCCTGCAGATAACCCGACCCCTGGTGGCTGCCCGAGCCTTTGTGCAGGGCTGGAGACTGGAAG AAATGTGGTCAGCGAAGCGCTTAAGGTTCCCGTGTCTGAAGGCTGCAGCCAGGCTCTGATGCT CTCATCGGCTGTCCCCTGTGCCGGGGGTCCCCTCACTTATGCCCTGCCAGGGCTTCTGCCTCA ACGTGGTTCGTGGCTGTCTCAGCAGCAGGGGACTGGAGCCTGACTGGGGCACTATCTGGATGG TCTCCTGATCCTGGCTGATAAGCTCCAGGGCCCTTTTCTTTGAGCTGACGGCCGAGTCCATT GGGGTGAAGATCTCGAGGGTTTGATGTACCTGCAGGAAAAAGTGCAGAGGTGTCCGCCAGG TATTTCAAGAGTGCGGCCCGCCCGACCCGGTGCTGCCCGCAACCGTCGAGCCCGCGCCCGC GGAAGAGGCGGGCCGGCTGTGGTGCATGGTGACCGAGGAGGAGCGGCCAACGACCGCCGAGGC ACCAACCTGCACCGGCTGGTGTGGGAGCTCCGCGAGCGTCTGGCCGGATGCGGGGCTTCTGGG CCCGGCTGTCCCTGACGGTGTGCGGAGACTCTCGATGGCAGCGGACGCTCTGGTGGAGCGGC GCCCTGCTGGACCGGAGCCGGGCGGGGCGGTACTTGCCGCACTGGTCCGGGGCTCCCCGGCC GAGCAGGTCAACAACCCGAGCTCAAGGTGGACGCTCGGGCCCGATGTCCGCACACGGCGGC GTCGGCTACAGCTCCGGGCGGCCACGGCCAGAATGAAACGGCCCACTGGGACACGACCTGGA CGGGCAGGACGAGATGAGGATGCCAGCGGCTCTGGAGGGGACAGCAGTATGCAGATGACTGG ATGGCTGGGGCTGTGGCTCCCCAGCCCGGCTCTCTCGGCTCCATACCTCTAGAGGGATG GTTCTGGGGGCAAGGAGGAGGTGGCAGTGCCGCTACAACAGGGCCGAGCAGGAGTGGGG GGCATCTATTGTTTTACACCAAAACCATCTCATTCTCTCCCTCTCAGCCCTGGCCCTGCTT GGACCTCGATAACGGGGGAGGGGTGCCCTAGCATCAGAAGGGTTATGCGCCTTTC		
	ORF Start: ATG at 1		ORF Stop: TAA at 1738

5

	SEQ ID NO: 62	579 aa	MW at 62828.7kD
NOV15b, CG50970-01 Protein Sequence	MSALRPLLLLLLPLCPGPGPGPGSEAKVTRSCAETROVLGARGYSLNLIIPPALISGEHLRVCPQ EYTCSSSETEQRLIRETEATFRGLVEDSGSFLVHTLAARHRKDFEFLEMLSVAOHSLTQLFSH SYGRLYAQHALIFNGLFSRLRDFYGESGEGLDDTLADFQAQLLERVFPLLPQYSPFPDYLLCL SRLASSTDGSLQFPFGDSRRLRLQITRTLVAARAFVQLETGRNVVSEALKVPVSEGCQALMR LIGCPLCRGVPSPMLPCQGFCLNVVRGCLSSRGLPEPDWGNLYLDGLLILADKLQGFPSFELTAESI GVKISEGLMYLQENSASAKVSAQVFQECGPPDPVPARNRRAPPPREAGRLWGMVTEERPTTAAG TNLHRLVWELRERLARMRGFWARLSLTVCGDSRMAADASLEAAPCWTGAGRGRYLPVVGGS PAEQVNNPELKVDAAGPDVPTRRRLQLRAATARMKTAALGHDLDGQDAEDAGSGGGGQYADWD MAGAVAPPARPPRPYPYPRRDGSGGKGGGGSARYNQGRSRSGGASIGFHTQTILILSLALALL GPR		

	SEQ ID NO: 63	1648 bp	
NOV15d, 274054257 DNA Sequence	CACCGGATCCAGCGAGGCAAGGTCAACCGGAGTTGTGCAGAGACCCGCGAGGTGCTGGGGGCC CGGGGATATAGCTTAAACCTAATCCCTCCCGCCCTGATCTCAGGTGAGCACCTCCGGGTCTGTC CCCAGGAGTACACCTGCTGTTCAGTGAGACAGAGCAGAGGCTGATCAGGGAGACTGAGGCCAC CTTCCGAGGCCTGGTGGAGGACAGCGGCTCCTTTCTGGTTCACACACTGGCTGCCAGGCACAGA AAATTTGATGAGTTTTTCTGGAGATGCTCTCAGTAGCCCACTCTCTGACCCAGCTCTTCT CCCCTCTACCGCCGCTGTATGCCAGCAGCCCTCATATTCAATGGCTGTTCTCTCGGCT GCGAGACTTCTATGGGAATCTGGTGAGGGTTGGATGACACCTGGCGGATTTCTGGGCACAG CTCCTGGAGAGAGTGTCCCGCTGCTGCACCCACAGTACAGCTTCCCCCTGACTACCTGCTCT GCCTCTCAGCCTTGGCTCATCTACCGATGGCTCTCTGCAGCCCTTGGGGACTCACCCGCGC CCTCCGCTGCAGATAACCCGACCCCTGGTGGCTGCCCGAGCCTTTGTGCAGGGCCTGGAGACT GGAAGAAATGTGGTCAGCGAAGCGCTTAAGGTGCCGGTGTCTGAAGGCTGCAGCCAGGCTCTGA TGCGTCTCATCGGCTGTCCCTGTGCCGGGGGTCCCCTCACTTATGCCCTGCCAGGCTTCTG		

	SEQ ID NO: 64	549 aa	MW at 59802.9kD
NOV15d, 274054257 Protein Sequence	TGSSEAKVTRSCAETROVLGARGYSINLIIPPALISGEHLRVCPQEYTCSSSETEQRLIRETEAT FRGLVEDSGSFLVHTLAARHKLFDEFFLEMLSVAQHSLTQLFHSYGRLYAQHALIFNGLFSRL RDFYGESGEGLLDTLADFWAQLLERVFPLLHPQYSFPDPYLLCLSRLASSTDGSLLQPFQDSPRR LRLQITRTLVAARAFVQGLEGTGRNVVSEALKVPVSEGCCSQUALMRLIGSCPLCRGVPVSLMPCQGC LNVVRGCLSSRGLEPDWGNLYLDGLLILADKLQGPFSFELTAESIGVKISEGLMYLQENSAKUSA QVFQECGPPDPVPARNRRAPPPREEAGRLWSMVTEERPTTAAGTNLHRLVWELRERLARMRGF WARLSLTVCGDSRMAADASLEAAPCWTGAGRGRYLPVVVGSSPAEQVNNPELKVDASGPDVPT RRRLQLRAATARMKTAALGHDLGDQDADEASGSGGQYADDWMAGAVAPPARPPRPYPYPPR DGSGGKGGGSARYNQGRSRSGGASIGFHTQTILLEG		

NOV15e, CG50970-03 DNA Sequence	SEQ ID NO: 65	1613 bp		
	ATGTCCGCGCTGCGACCTCTCCTGCTTCTGCTGCTCCTGTGCCGTCTCGGTCCCAGGAC CCGGAGCGAGGCAAAGGTCAACCGGAGTTGTGCAGAGACCCGGCAGGTGCTGGGGGCCCGGGG ATATAGCTTAAACCTAATCCCTCCGCCCCTGATCTCAGGTGAGCACCTCGGGGTCTGTCCCAG GAGTACACCTGCTGTCTCCAGTGAGACAGAGCAGAGGCTGATCAGGGAGACTGAGGCCACCTTCC GAGGCTGGTGGAGGACAGCGGCTCCTTTCTGGTTTCAACACTTGCTGGTCGAGGCACAGAAAATT TGATGAGTTTTTTCTGAGATGCTCTCAGTAGCCAGCACTCTGAGCCAGCTCTTCTCCCAC TCCTACGGCCGCTGTATGCCAGCAGCGCCCTCATATTCAATGGCCTGTTCTCTCGGCTGCGAG ACTTCTATGGGGAATCTGGTGAGGGGTGGATGACACCTGGCGGATTCTGGGCAAGCTCCT GGAGAGAGTGTCCCGCTGCACCCAGTACAGCTTCCCCCTGACTACCTGCTGCTCTC TCACGCTTGGCCTCATCTACCGATGGCTCTCTGAGCCCTTTGGGGACTCACCGCCGCTCTCC GCCTGCAGATAACCGGACCCCTGGTGGCTGCCGAGCCTTTGTGTCAGGGCCTGGAGACTGGAAG AAATGTGGTCAGCGAAGCGCTTAAGGTGCCGGTGTCTGAAGGCTGCAGCCAGGCTCTGATGCGT CTATCGGCTGTCCCTGTGTCGCGGGGGTCCCTCACTTAGCCCTGCCAGGGCTCTGCTCA ACGTGGTTCGTGGCTGTCTCAGCAGCAGGGGACTGGAGCCTGACTGGGGCACTATCTGGATGG TCTCTGTATCTGGCTGATAAGCTCCAGGGCCCTTTTCCTTTGAGCTGACGGCCGAGTCCATT GGGGTGAAGATCTCGGAGGGTTTGATGTACCTGCAGGAAACAGTGCAGAGGTGTCGCGCCAGG TGTTTCAGGAGTGCAGGCCCCCGACCCGGTGCTGCCGCGACCGTGCAGCCCCGCGCCCCG GGAAGAGGGCGGCGGCTGTGGTGCATGGTGACCGAGGAGGAGCGGCCACGACGGCCGAGGC ACCAACCCTGCACCGGCTGGTACTTGCCGCCAGTGGTCCGGGGCTCCCCGGCCGAGCAGGTCAAC AACCCCGAGCTCAAGGTGAGCGCTCGGGCCCCGATGTCCCCGACGGCGGGCTCGGCTACAGC TCCGGGCGGCCACGGCCAGAATGAAACGGCCGCACTGGGACAGCACTGGACGGCAGGACGC GGATGAGGATGCCAGCGGCTCTGAGGGGGACAGCAGTATGCAGATGACTGGATGGCTGGGGCT GTGGCTCCCCAGCCCGGCTCTCTCGGCTCCATACCCCTCTAGAGGGATGGTTCTGGGGCA AAGGAGGAGGTGGCAGTGGCCGCTCAACACGGGCCGAGCAGGAGTGGGGGGCATCTATTGG TTTTACACCCAAACCATCCTCATTCTCTCCCTCTCAGACCTGGCCCTGCTTGGACCTCGATAA CGGGGGAGGGGTG			
	ORF Start: ATG at 1		ORF Stop: TGA at 1348	

177

CG50970-03 Protein Sequence	EYTCCSSETEQRLIRETEATFRGLVEDSGSFLVHTLAARHRKFDEFFLEMLSVAQHSLTQLFSH SYGRLYAQHALIFNGLFSRLRDFYGESGEGLDDTLADFWAQLLERVFPPLHPOYSFPPDYLLCL SRLASSTDGSLQPFQDSPRRLRLQITRTLVAARAFVQGLETGRNVVSEALKVPVSEGCQSALMR LIGCPLCRGVPPLMPCQGFCLNVVRGCLSSRGLEPDWGNLYLDGLLILADKLQGPFSFELTAESI GVKISEGLMYLQENSAKVSQVFOECGPPDPVPARNRRAPPPREEAGRLWSMVTEERPTTAAG TNLHRLVLAASGRGLPGRAGQPPRAQGGRLGPRCPDTAASATAPGGHGQENGRGTGTRPGRAGR G
--------------------------------	--

	SEQ ID NO: 67	1297 bp	
NOV15f, 237922026 DNA Sequence	CACCGGATCCACCAGCGAGGCAAAGGTCACCCGGAGTTGTGCAGAGACCCGGCAGGTGCTGGGG GCCCGGGGATATAGCTTAAACCTAATCCCTCCCGCCCTGATCTCAGGTGAGCACCTCCGGGTCT GTCCCCAGGAGTACACCTGCTGTTCCAGTGAGACAGAGCAGAGGCTGATCAGGAGACTGAGGC CACCTCCGAGGCCTGGTGGAGGACAGCGGCTCCTTTCTGGTTACACACTGGCTGCCAGGCAC AGAAAATTTGATGAGTTTTTCTGGAGATGCTCTCAGTAGCCAGCACTCTCTGACCCAGCTCT TCTCCCACTCTACGGCCGCTGTATGCCAGCAGCCCTCATATTCAATGGCCTGTTCTCTCG GCTGCGAGACTTCTATGGGGAATCTGGTGAGGGGTTGGATGACACCTGGCGGATTTCTGGGCA CAGCTCCTGGAGAGAGTGTCCCGCTGCTGCAACCCACAGTACAGCTTCCCCCTGACTACCTGC TCTGCCTCTCAGCTTGGCCTCATCTACCGATGGCTCTCTGCAGCCCTTTGGGGACTCACCCCG CCGCCTCCGCCTGCAGATAACCCGGACCTGGTGGCTGCCGAGCCTTTGTGCAGGGCCTGGAG ACTGGAAGAAATGTGGTCAGCGAAGCGCTTAAGGTGCCGGTGTCTGAAGGCTGCAGCCAGGCTC TGATGCGTCTCATCGGCTGTCCCTGTGCCGGGGGTCCTTACTTATGCCCTGCCAGGGCTT CTGCTCAACGTGGTTCTGGTGTCTCAGCAGCAGGGGACTGGAGCTGACTGGGGCAACTAT CTGGATGGTCTCCTGATCCTGGCTGATAAGCTCCAGGGCCCCCTTTCTTTGAGCTGACGGCCG AGTCCATTGGGGTGAAGATCTCGGAGGGTTGATGTACCTGCAGGAAACAGTGCGAAGGTGTC CGCCAGGTGTTTCAGGAGTGCGGCCCCCGACCCGGTGCCTGCCCGCAACCGTCGAGCCCCG CCGCCCCGGGAAGAGGCGGGCCGCTGTGGTTCGATGGTGACCGAGGAGGAGCGGCCACGACGG CCGCAGGCACCAACCTGCAACGGCTGGTACTTGCCGCCAGTGGTCGGGGCTCCCCGGCCGAGC AGGTCAACAACCCGAGCTCAAGGTGGACGCTCGGGCCCCGATGTCCTGCACGGCGGCGTCTG GCTACAGCTCCGGGCGGCCACGGCCAGAATGAAACGGCCGCACTGGGACAGACCTGGACGGG CAGGACGCGGACTCGAG		
	ORF Start: at 2	ORF Stop: end of sequence	

5

	SEQ ID NO: 68	432 aa	MW at 47040.8kD
NOV15f, 237922026 Protein Sequence	TGSTSEAKVTRSCAETROVLGARGYSNLNLIIPALISGEHLRVCPQEYTCSSSETEQRLIRETEA TFRGLVEDSGSFLVHTLAARHRKFDEFFLEMLSVAQHSLTQLFSHSYGRLYAQHALIFNGLFSR LRDFYGESGEGLDDTLADFWAQLLERVFPPLHPOYSFPPDYLLCLSRASSTDGSLQPFQDSPR RLRLQITRTLVAARAFVQGLETGRNVVSEALKVPVSEGCQSALMRIGCPLCRGVPPLMPCQGF CLNVVRGCLSSRGLEPDWGNLYLDGLLILADKLQGPFSFELTAESI GVKISEGLMYLQENSAKVS AQVFOECGPPDPVPARNRRAPPPREEAGRLWSMVTEERPTTAAGTNLHRLVLAASGRGLPGR AQQPRAQGGRLGPRCPDTAASATAPGGHGQENGRGTGTRPGRAGRGL		

	SEQ ID NO: 69	1126 bp
NOV15g, 237922511 DNA Sequence	CACC GGATCC ACCAGCGAGGCAAAGGTCA CCGGAGTTGTGCAGAGACCCGGCAGGTGCTGGGG GCCCCGGGATATAGCTTAAACCTAATCCCTCCCGCCCTGATCTCAGGTGAGCACTCCGGGTCT GTCCCCAGGAGTACACCTGCTGTTCCAGTGAGACAGAGCAGAGGCTGATCAGGGAGACTGAGGC CACCTTCCGAGGCCTGGTGGAGGACAGCGGCTCCTTTCTGGTTACACACTGGCTGCCAGGCAC AGAAAAATTTGATGAGTTTTTCTGGAGATGCTCTCAGTAGCCAGCACTCTCTGACCCAGCTCT TCTCCCACTCCTACGGCCGCTGTATGCCAGCAGCCCTCATATTCAATGGCCTGTTCTCTCG GCTGCGAGACTTCTATGGGGAATCTGGTGAGGGGTTGGATGACACCCCTGGCGGATTTCTGGGCA CAGTCTCTGGAGAGAGTGTCCCGCTGCTGCACCCACAGTACAGCTTCCCCCTGACTACCTGC TCTGCCCTCTCAGCTTGGCCTCATCTACCGATGGCTCTCTGCAGCCCTTGGGGACTCACCCCG CCGCTCCGCTGCAGATAACCCGACCCCTGGTGGCTGCCGAGCCTTTGTGCAGGGCCTGGAG ACTGGAAGAAATGTGGTCAGCGAAGCGCTTAAGGTGCCGGTGTCTGAAGGCTGCAGCCAGGCTC TGATGCGTCTCATCGGCTGTCCCTGTGCCGGGGGTCCCTCACTTATGCCCTGCCAGGGCTT CTGCCTCAACGTGGTTCGTGGCTGTCTCAGCAGCAGGGGACTGGAGCCTGACTGGGGCACTAT CTGGATGGTCTCTGATCCTGGCTGATAAGCTCCAGGGCCCCCTTTTCTTGAAGCTGACGGCCG AGTCCATTGGGGTGAAGATCTCGAGGGTTTGTATGTACCTGCAGGAAAACAGTGCAGAGGTGTC CGCCAGGTGTTTACAGAGTGCAGCCCCCCCCGACCCGGTGCCTCCGCAACCGTCGAGCCCCG CCGCCCCGGGAAGAGCGGGCCGGCTGTGGTGTGATGGTGACCGAGGAGGAGCGGCCACGACGG CCGAGGCACCAACCTGCACCGCTGGTACTTCTCGAG	
	ORF Start: at 2	ORF Stop: end of sequence

	SEQ ID NO: 70	375 aa	MW at 41526.8kD
NOV15g, 237922511 Protein Sequence	TGSTSEAKVTRSCAETRQVLGARGYSNLNIPALISGEHLRVCPQBYTCCSSETEQRLIRETEA TFRGLVEDSGSFLVHTLAARHRKFDEFFLEMLSVAQHSLTQLFHSYGRLYAQHALIFNGLFSR LRDFYGESGEGLDLTLADFWAQLLERVFLHPQYSPDPDYLLCLSRLASSTDGSLOPFGDSPR RLRLQITRTLVAARAFVQLEETGRNVVSEALKVPVSEGCQALMRLIGCPLCRGVPSLMPCCGF CLNVVRGCLSSRGLPEPDWGNLYLDGLLILADKLQGPFSFELTAESIGVKISEGLMYLQENSAKVS AQVFQECGPPDPVPARNRRAPPPREEAGRLWSMVTEERPTTAAGTNLHRLVLE		

5

	SEQ ID NO: 71	1776 bp
NOV15h, 315490136 DNA Sequence	CACC GGATCC ACCATGTCCCGCTGCGACCTCTCCTGCTCTGCTGCTGCTCTGCTCCGGT CCTGGTCCCGGACCCGGGAGCGAGCAAAGGTCA CCGGAGTTGTGCAGAGACCCGGCAGGTGCT TGGGGGCCCGGGGATATAGCTTAAACCTAATCCCTCCCGCCCTGATCTCAGGTGAGCACTCCG GGTCTGTCCCCAGGAGTACACCTGCTGTTCCAGTGAGACAGAGCAGAGGCTGATCAGGGAGACT GAGGCCACCTTCCGAGGCCTGGTGGAGGACAGCGGCTCCTTTCTGGTTACACACTGGCTGCCA GGCACAGAAAATTTGATGAGTTTTTCTGGAGATGCTCTCAGTAGCCAGCACTCTGACCCCA GCTCTTCTCCCACTCCTACGCGCCCTGTATGCCAGCAGCCCTCATATTCAATGGCCTGTTCT TCTCGGCTGCGAGACTTCTATGGGGAATCTGGTGAGGGGTTGGATGACACCCCTGGCGGATTTCT GGGCACAGCTCCTGGAGAGAGTGTCCCGCTGCTGCACCCACAGTACAGCTTCCCCCTGACTA CCTGCTCTGCCTCTCAGCTTGGCCTCATCTACCGATGGCTCTCTGCAGCCCTTGGGGACTCA CCCCGCCGCTCCGCTGCAGATAACCCGACCCCTGGTGGCTGCCGAGCCTTTGTGCAGGGCC TGGAGACTGGAAGAAATGTGGTCAGCGAAGCGCTTAAGGTGCCGGTGTCTGAAGGCTGCAGCCA GGCTCTGATGCGTCTCATCGGCTGTCCCTGTGCCGGGGGTCCCTCACTTATGCCCTGCCAG GGCTTCTGCCTCAACGTGGTTCTGGCTGTCTCAGCAGCAGGGGACTGGAGCCTGACTGGGGCA ACTATCTGGATGGTCTCCTGATCCTGGCTGATAAGCTCCAGGGCCCCCTTTCTTGTGAGCTGAC GGCCGAGTCCATTGGGGTGAAGATCTCGAGGGTTTGTATGTACCTGCAGGAAAACAGTGCAGAG GTGTCCGCCAGGTGTTTTCAGGAGTGCAGCCCCCGGACCCGGTGCCTGCCCGCAACCGTGCAG CCCCGCCGCCCGGGGAGAGGCGGGCCGGCTGTGGTGCATGGTGACCGAGGAGGAGCGGCCAC GACGGCCGCGAGGCACCAACCTGCACCGGCTGGTGTGGAGCTCCGCGAGCGTCTGGCCCGGAT CGGGGCTTCTGGGCGCGCTGTCCCTGACGGTGTGCGGAGACTCTCGCATGGCAGCGGACGCT CGCTGGAGGCGCGCCCTGTGGAACCGAGCCGGGCGGGGCGGTACTTCCCGCCAGTGGTCCG GGGCTCCCCGGCCGAGCAGGTCAACAACCCGAGCTCAAGGTGGAAGCCTCGGGCCCCGATGTC CCGACACGGCGGCGTCACTACAGCTCCGGGCGGCCACGGCCAGAAATGAAAACGGCCGACACTG GACACGACCTGGACGGGCGAGGACGGGATGAGGATGCCAGCGCTCTGGAGGGGACAGCAGTA TGCAGATGACTGGATGGCTGGGGCTGTGGCTCCCCAGCCCGCTCTCGGCTCTCATACCT CCTAGAAGGGATGGTTCTGGGGCAAAGGAGGAGTGGCAGTCCCGCTACAACAGGGCCGGA GCAGGAGTGGGGGGCATCTATTGGTTTTCAACCCAAACCATCTCTCTCTCTCTCTCAGC CCTGGCCCTGCTTGGACCTCGACTCGAGGGCAAGGGCAATTCCAGCA	
	ORF Start: at 2	ORF Stop: end of sequence

510180

	SEQ ID NO: 75	725 bp	
NOV15j, CG50970-04 DNA Sequence	CGCCTGGTCCAGCTATCGTGCTCGGTATTTCAGTTTCCGGAGCAGCGCTCTTTCTCTGGCCCCG GGAACGGTCCCGCGGCCGAGTACCGGATTCCCGAGTTTGGGAGGCTCTGCTTCTCCTTAGGA CCCACTTGCCGCTCCTGGGGTGGCTGCAGTTATGTCGCGCTGCGACCTCTCCTGCTTCTGCTG CTGCTCTGTGTCCCGGTCTGGTCCCGACCCGGGAGCGAGGCAAGGTCAACCGGAGTTGTG CAGAGACCCGGCAGGTGCTGGGGGCCCGGGGATATAGCTTAAACCTAATCCCTCCCGCCCTGAT CTCAGGTGAGCACCTCCGGTCTGTCCCAGGAGTACACCTGCTGTTCAGTGAGACAGAGCAG AGGCTGATCAGGGAGACTGAGGCCACCTTCCGAGGCCTGGTGGAGGACAGCGGCTCTTTCTG TTCAACACTGGCTGCCAGGCAAGAAAATTTGATGAGTTTTCCTGGAGATGCTCTCAGTAGC CCGGCTCTCGGCTCCATACCTCTCTAGAAGGGATGGTTCTGGGGGCAAGGAGGAGGTGGC AGTGCCCGCTACAACAGGGCCGGAGCAGGAGTGGGGGGCATCTATTGGTTTTCACACCCAAA CCATCCTCATTCTCTCCCTCTCAGCCCTGGCCTTGCTTGGACCTCGATAACCGGGGAGGGGTGC CCTAGCATCAGAAGGGTTCAT		
	ORF Start: ATG at 160		ORF Stop: TAA at 688

	SEQ ID NO: 76	176 aa	MW at 18879.4kD
NOV15j, CG50970-04 Protein Sequence	MSALRPLLLLLLPLCPGPGPGSGEAKVTRSCAETRVQLGARGYSNLIPPALISGEHLRVCPQ EYTCCSSETEQRLIRETEATFRGLVEDSGSFLVHTLAARHRKFDEFFLEMLSVARPPRPPYPPR RDGSGGKGGGGSARYNQGRSRSGASIGFHTQTLLILSLALALLGPR		

	SEQ ID NO: 77	1590 bp	
NOV15k, CG50970-05 DNA Sequence	AGCGAGGCAAGGTCAACCGGAGTTGTGCAGAGACCCGGCAGGTCTGGGGGCCCGGGGATATA GCTTAAACCTAATCCCTCCCGCCCTGATCTCAGGTGAGCACCTCCGGGTCTGTCCCAGGAGTA CACTGTCTGTTCCAGTGAGACAGAGCAGAGGCTGATCAGGGAGACTGAGGCCACCTTCCGAGGC CTGGTGGAGGACAGCGGCTCTTTCTGGTTACACACTGGCTGCCAGGCACAGAAAATTTGATG AGTTTTTCTGGAGATGCTCTCAGTAGCCCAACACTCTCTGACCCAGCTCTTCTCCCACTCCTA CGGCCGCTGTATGCCAGCACGCCCTCATATTCAATGGCCTGTTCTCTCGGCTGCCAGACTTC TATGGGGAATCTGGTGAGGGGTTGGATGACACCTGGCGGATTTCTGGGCACAGCTCCTGGAGA GAGTGTTCCTGCTGTGCACCCACAGTACAGCTTCCCCCTGACTACCTGCTCTGCCTCTCACG CTTGGCCTCATCTACCGATGGCTCTCTGCAGCCCTTGGGGACTCAACCGCGCCCTCCGCTG CAGATAACCCGGACCTGGTGGCTGCCGAGCCTTTGTGCAGGGCTGGAGACTGGAAGAAATG TGGTCAGCGAAGCGCTTAAGGTGCCGGTGTCTGAAGGCTGCAGCCAGGCTCTGATGCGTCTCAT CGGCTGTCCCTGTGCCGGGGGGTCCCTCCTATATGCCCTGCCAGGGCTTCTGCTCAACGTG GTTCGTGGCTGTCTCAGCAGCAGGGGACTGGAGCCTGACTGGGGCACTATCTGGATGGTCTCC TGATCCTGGCTGATAAGCTCAGGGCCCTTTTCTTTGAGCTGACGGCCGAGTCCATTGGGGT GAAGATCTCGGAGGGTTTGATGTACCTGCAGGAAAACAGTGCGAAGGTGTCCGCCAGGTGTTT CAGGAGTGCGGCCCCCGACCCGGTGCCTGCCGCAACCGTCGAGCCCCCGCGCCCGGGAAG AGGCGGGCCCGCTGTGGTGCATGGTGACCGAGGAGCGGCCACAGCGCCCGAGGCCACCAA CCTGCACCGCTGGTGTGGGAGCTCCGCGAGCGTCTGGCCCGGATGCGGGGCTTCTGGGCCCGG CTGTCCCTGACGGTGTGCCGAGACTCTCGCATGGCAGCGGACGCTCGCTGGAGGACGCGCCT GCTGGACCGGAGCCGGGCGGGGCGGTACTTGCCGCCAGTGGTGGGGGCTCCCGGCCGAGCA GGTCAACAACCCCGAGCTCAACGTGGACGCTCGGGCCCCGATGTCCCGACAGCGCGCGTGG CTACGGCTCCGGGCGGCCAGGCCAGATGAAAACGGCCGCACTGGGACAGCACTGGACGGGC AGGACGCGGATGAGGATGCCAGCGGCTCTGGAGGGGGAAGCAGATATGAGATGACTGGATGGC TGGGGCTGTGGCTCCCCAGCCCGGCTCTCGGCTCCATACCTCCTAGAAGGATGGTTCT GGGGCAAGGAGGAGGTGGCAGTGCCCGCTACAACAGGGCCGAGCAGGT		
	ORF Start: at 1		ORF Stop: end of sequence

10

	SEQ ID NO: 78	530 aa	MW at 57988.9kD
NOV15k, CG50970-05 Protein Sequence	SEAKVTRSCAETRVQLGARGYSNLIPPALISGEHLRVCPQ EYTCCSSETEQRLIRETEATFRG LVEDSGSFLVHTLAARHRKFDEFFLEMLSV AQHSLTQLFSHSYGRLYAQHALIFNGLFSRLRDF YGESGEGLDLADFWAQLERVFLLHPQYSFPDYLCLSLASSTDGSLQFPFGDSPRRLRL QITRTLVAARAFVQLETGRNVVSEALKVPVSEGCSSQALMRLIGPLCRGVPSPMPCQGFCLNV VRGCLSSRGLEPDWGNLYDLGLILADLKQGPFSFELTAESIGVKISEGLMYLQENSARKVSAQVF QECGPPDPVPARNRRAPPPREEAGRLWSMVTEERPTTAAGTNLHRLVWELERLARMRGFWAR LSLTVCGDSRMAADASLEAPCWTGAGRGRLPFPVVGSPAEQVNNPELVNDASGPDVPTRRRR		

	LRLRAATARMKTAALGHDLGQDAEDASGSGGGQQYADDWMAGAVAPPARPPRPPYPPRDSGSGKGGGGGSARYNQGRSRS
--	--

	SEQ ID NO: 79	1762 bp
NOV151, CG50970-07 DNA Sequence	<p> CACC GGAT ACCACCATGTCCGCGCTGCGACCTCTCCTGCTTCTGCTGCTGCCTCTGTGTCCCGGT  CCTGGTCCCGGACCCGGGAGCGAGGCAAAGGTCAACCCGGAGTTGTGCAGAGACCCGGCAGGTGC  TGGGGGCCCCGGGGATATAGCTTTAAACCTAATCCCTCCCGCTGATCTCAGGTGAGACCTCCG  GGTCTGTCCCGAGGAGTACACCTGCTGTTCAGTGAGACAGAGCAGAGGCTGATCAGGGAGACT  GAGGCCACCTTCCGAGCGCTGGTGGAGGACAGCGGCTCCTTTCTGGTTTACACACTGGCTGCCA  GGCACAGAAAATTGATGAGTTTTTCTGGAGATGCTCTCAGTAGCCCAAGCACTCTCTGACCCA  GCTCTTCTCCCACTCTACGGCGCCCTGTATGCCAGCAGCCCTCATATTCAATGGCCTGTTTCT  TCTCGGCTGCGAGACTTCTATGGGGAATCTGGTGAGGGGTTGGATGACACCTTGGCGGATTTCT  GGGCA CAGCTCTGGAGAGAGTGTTCGCGCTGCTGCACCCACAGTACAGCTTCCCCCTGACTA  CTGCTCTGCTCTCAAGCTTGGCCTCATCTACCGATGGCTCTCTGCAGCCCTTTGGGGACTCA  CCCCGCGCCTCCGCTTCGAGATAAACCCGGAACCTGGTGCTGCCCCAGCCTTTGTGCAGGGCC  TGGAGACTGGAAGAAATGTGGTCAGCGAAGCGCTTAAGGTGCCGGTGTCTGAAGGCTGCAGCCA  GGCTCTGATGCGCTCATACGGCTGTCCCTGTGCGGGGGGTCCCCTCACTTATGCCCTGCCAG  GGCTTCTGCCTCAAGCTGGTTCTGTGGCTGTCTCAGCAGCAGGGGACTGGAGCCCTGACTGGGGCA  ACTATCTGGATGGTCTCCTGATCCTGGCTGATAAGCTCCAGGCCCCCTTTTCTTTGAGCTGAC  GGCCGAGTCCATTGGGGTGAAGATCTCGGAGGGTTTGATGTACTTGCAGGAAAACAGTGCGAAG  GTGTCGCGCCAGGTGTTTCAGGAGTGCGGCCCCCCCCGACCCGGTGCTGCCCCGAACCGTCGAG  CCCCGCGCCCCGGGAAGAGGCGGGCTGTGGTGCATGGTGACCGAGGAGGAGCGGGCCAC  GACGGCCGCGAGGCACCAACCTGCA CCGGCTGGTGTGGGAGACTCTCGCATGGCAGCGGACGCC  CGGGGCTTCTGGGCCCCGGCTGTCCCTGACGSGTGTGCGGAGACTCTCGCATGGCAGCGGACGCC  CGCTGGAGGCGCGCCCTGTGGA CCGGAGCCGGGCGGGGCGGGTACTTGCGCCAGTGGTCCG  GGCTCCCCCGCCGAGCAGGTCAACAAACCCGAGCTCAAGGTGGACGCCCTCGGGCCCCGATGTC  CCGACACGGCGGCGCTCGGCTACAGCTCCGGGCGGCCACGCCCAAGTGA AAAACGGCCGCACTGC  GACACGACTGGACGGG CAGGACGCGGATGAGGATGCCAGCGGCTCTGGAGGGGGACAGCAGTA  TG CAGATGATCGGATGGCTGGGGCTGTGGCTCCCCAGCCCGGCTCTCTCGCCTCCATACCTT  CCTAGAAAGGATGGTTCTGGGGGCAAAGGAGGAGTGCGCCGCTACAACAGGGGCCGGA  GCAGGAGTGGGGGGCATCTATTGGTTTTTCAACCCAAACATCCTCATTCTCTCTCAGC  CCTGGCCCTGCTTGGACCTCGATAGGTGACGGC </p>	
	ORF Start: ATG at 14	ORF Stop: TAG at 1751

5

	SEQ ID NO: 80	579 aa	MW at 62828.7kD
NOV151, CG50970-07 Protein Sequence	MSALRPILLLLPLCPGPGPGPGSEAKVTRSCAETROVLGARGYSNLNIPPALISGEHLRVCPQ EYTCSSSETEQRLIRETEATFRGLVEDSGSFLVHTLAARHKRQDEFFLEMLSVAQHSLTQLFSH SYGRLYAQHALIFNGLFSRLRDFYGESGEGLDLDTLADFWAQLLRFVFLMLHPQYSFFPDYLLCL SRLASSTDGSLLQPPGDSRRRLRLQITRTLVAARAFVQGLETRGNVSEALKVPVSEGCQSALMR LIGCPLCRGVPSLMPCCQGFCLNVVRGCLSSRGLEPDWGNLYLDGLLILADKLQGPFFELTAESI GVKISEGLMYLQENSARKVSAQVFQECGPPDPVPARNRRAPPPREEAGRLWSMVTEERPTTAAG TNLHRLVWELRERLARMRGFWARLSLTVCGDSRMAADASLEAAPCWTGAGRGRYLPPVVGGS PAEQVNNPELKVDAAGSDVPTRRRLQLRAATARMKTAALGHDLDGQDADESDSGSGGQYADDW MAGAVAPPARPPRPYPFRDGGSGGKGGGSARYNQGRSRSGGASIGFHTQTILILSLALALL GPR		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 15B.

10

Table 15B. Comparison of NOV15a against NOV15b through NOV15l.		
Protein Sequence	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region

NOV15b	1..579 1..579	579/579 (100%) 579/579 (100%)
NOV15c	1..579 1..579	579/579 (100%) 579/579 (100%)
NOV15d	24..567 4..547	543/544 (99%) 544/544 (99%)
NOV15e	1..391 1..391	391/391 (100%) 391/391 (100%)
NOV15f	21..391 2..372	369/371 (99%) 369/371 (99%)
NOV15g	21..391 2..372	369/371 (99%) 369/371 (99%)
NOV15h	1..579 5..583	579/579 (100%) 579/579 (100%)
NOV15i	1..380 1..380	379/380 (99%) 380/380 (99%)
NOV15j	1..119 1..119	118/119 (99%) 119/119 (99%)
NOV15k	24..553 1..530	528/530 (99%) 529/530 (99%)
NOV15l	1..579 1..579	579/579 (100%) 579/579 (100%)

Further analysis of the NOV15a protein yielded the following properties shown in Table 15C.

Table 15C. Protein Sequence Properties NOV15a	
SignalP analysis:	Cleavage site between residues 24 and 25
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 5; pos.chg 1; neg.chg 0  H-region: length 19; peak value 10.14  PSG score: 5.74</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): -0.46  possible cleavage site: between 20 and 21</p> <p>&gt;&gt;&gt; Seems to have a cleavable signal peptide (1 to 20)</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 21  Tentative number of TMS(s) for the threshold 0.5:  1  Number of TMS(s) for threshold 0.5: 0</p>

	<p>PERIPHERAL Likelihood = 4.24 (at 254)  ALOM score: -0.85 (number of TMSs: 0)</p> <p>MTOP: Prediction of membrane topology (Hartmann et al.)  Center position for calculation: 10  Charge difference: -1.0 C( 1.0) - N( 2.0)  N &gt;= C: N-terminal side will be inside</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 1 Hyd Moment (75): 8.46  Hyd Moment (95): 7.50 G content: 4  D/E content: 1 S/T content: 2  Score: -4.90</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  R-3 motif at 45 ARGY S</p> <p>NUCDISC: discrimination of nuclear localization signals  pat4: RHRK (3) at 103  pat4: RRRR (5) at 468  pat7: PRRLRLQ (5) at 210  pat7: PARNRRA (4) at 353  pat7: PTRRRRL (5) at 466  bipartite: none  content of basic residues: 10.9%  NLS Score: 1.23</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals:  XXRR-like motif in the N-terminus: SALR</p> <p>none</p> <p>SKL: peroxisomal targeting signal in the C-terminus:  none</p> <p>PTS2: 2nd peroxisomal targeting signal: none</p> <p>VAC: possible vacuolar targeting motif: none</p> <p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif:  type 1: none  type 2: none</p> <p>NMYR: N-myristoylation pattern : none</p> <p>Prenylation motif: none</p> <p>memYQRL: transport motif from cell surface to Golgi:  none</p> <p>Tyrosines in the tail: none</p> <p>Dileucine motif in the tail: none</p> <p>checking 63 PROSITE DNA binding motifs: none</p>
--	--

	<p>checking 71 PROSITE ribosomal protein motifs: none</p> <p>checking 33 PROSITE prokaryotic DNA binding motifs: none</p> <p>NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination Indication: cytoplasmic Reliability: 55.5</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions total: 0 residues</p> <p>-----</p> <p>Final Results (k = 9/23):</p> <p>33.3 %: extracellular, including cell wall 33.3 %: mitochondrial 22.2 %: endoplasmic reticulum 11.1 %: vacuolar</p> <p>&gt;&gt; indication for CG50970-06 is exc (k=9)</p>
--	---

A search of the NOV15a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 15D.

5

Table 15D. Geneseq Results for NOV15a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG70277	Human Glypican-2 Precursor-like protein #1 - Homo sapiens, 579 aa. [WO200255702-A2, 18-JUL-2002]	1..579 1..579	579/579 (100%) 579/579 (100%)	0.0
ABG70279	Human Glypican-2 Precursor-like protein #3 - Homo sapiens, 449 aa. [WO200255702-A2, 18-JUL-2002]	1..391 1..391	391/391 (100%) 391/391 (100%)	0.0
ABG70278	Human Glypican-2 Precursor-like protein #2 - Homo sapiens, 465 aa. [WO200255702-A2, 18-JUL-2002]	1..380 1..380	378/380 (99%) 379/380 (99%)	0.0
AAU29071	Human PRO polypeptide sequence #48 - Homo sapiens, 555 aa. [WO200168848-A2, 20-SEP-2001]	2..511 7..504	227/512 (44%) 329/512 (63%)	e-127
AAB44256	Human PRO705 (UNQ369) protein	2..511 7..504	227/512 (44%) 329/512 (63%)	e-127

	sapiens, 555 aa. [WO200053756-A2, 14-SEP-2000]			
--	--	--	--	--

In a BLAST search of public sequence databases, the NOV15a protein was found to have homology to the proteins shown in the BLASTP data in Table 15E.

Table 15E. Public BLASTP Results for NOV15a				
Protein Accession Number	Protein/Organism/Length	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q8N158	Similar to cerebroglycan (Hypothetical protein FLJ38962) - Homo sapiens (Human), 579 aa.	1..579 1..579	579/579 (100%) 579/579 (100%)	0.0
P51653	Glypican-2 precursor (Cerebroglycan) (HSPG M13) - Rattus norvegicus (Rat), 579 aa.	1..579 1..579	477/581 (82%) 513/581 (88%)	0.0
Q9R087	Glypican-6 precursor - Mus musculus (Mouse), 555 aa.	2..511 7..504	227/512 (44%) 332/512 (64%)	e-127
Q9Y625	Glypican-6 precursor - Homo sapiens (Human), 555 aa.	2..511 7..504	227/512 (44%) 329/512 (63%)	e-127
Q8R3X6	Similar to glypican 6 - Mus musculus (Mouse), 565 aa.	2..511 7..514	228/522 (43%) 333/522 (63%)	e-125

5

PFam analysis indicates that the NOV15a protein contains the domains shown in the Table 15F.

Table 15F. Domain Analysis of NOV15a			
Pfam Domain	NOV15a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Glypican	3..566	271/631 (43%) 510/631 (81%)	6.7e-291

**Example 16.**

The NOV16 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 16A.

Table 16A. NOV16 Sequence Analysis			
	SEQ ID NO: 81	1242 bp	
NOV16a, CG54443-03 DNA Sequence	ATGGGGTCGACCGACTCCAAGCTGAACCTCCGGGAAGGCGGTGATCCAGCTCACCACCAAGACGCA GCCCGTGGGAAGCCACCGATGATGCCCTTTTGGGACCAGTTCTGGGCAGACACAGCCACCTCGGTGC AGGATGTGTTTGCACTGGTGCCGGCAGCAGAGATCCGGGCCGTGCGGGAAGAGTCAACCTCCAAC TTGGCCACCTGTGCTACAAGGCCGTTGAGAGGCTGGTGCAAGGAGCTGAGAGTGGCTGCCACTC GGAGAAGGAGAAGCAGATCGTCTGAACTGCAGCCGGCTGCTCAACCGCGTGTGCCCTACATCT TTGAGGACCCCGACTGGAGGGGCTTCTTCTGGTCCACAGTGCCCGGGCAGGGCAGGAGGGCAG GGAGAAGAGGATGATGAGCATGCCAGGCCCTGGCCGAGTCCCTGCTCCTGGCCATTGCTGACCT GCTCTTCTGCCCGGACTTCACGGTTTCAGAGCCACCGAGGAGCACTGTGGACTCGGCAGAGGACG TCCACTCCCTGGACAGCTGTGAATACATCTGGGAGGCTGGTGTGGGCTTCGCTCACTCCCCCAG CCTAACTACATCCAGATATGAACCGGATGGAGCTGCTGAAACTGCTGCTGACATGCTTCTCCGA GGCCATGTACCTGCCCCAGCTCCGGGAAAGTGGCAGCACCAACCATGGGTTCACTTCTTTGTT CCACGGAGAACAGACATGCCCTGCCCTCTTCACTCCCTCCTCAACACCGTGTGTGCTATGAC CCTGTGGGCTACGGGATCCCCACCAACCTGCTCTTCTGACTACCGGGAACCCCTGGTGGGA GGAGGCTGCCAGGTGCTCATTTGTCACTTTGGACCACGACAGTGCCAGCAGTGCCAGCCCCACTG TGGACGGCACCACCACTGGCACCAGCATGGATGATGCCGATCCTCCAGGCCCTGAGAACCTGTTT GTGAACCTACCTGTCCCGCATCCATCGTGAGGAGGACTTCCAGTTTCATCCTCAAGGGTATAGCCCG GCTGCTGTCAACCCCTGCTCCAGACCTACCTGCCTAACTCCACCAAGAAGATCCAGTTCCACC AGGAGCTGCTAGTTCTTCTGGAAGCTCTGCGACTTCAACAAGAAATTCCTTCTTCTGCTGCTG AAGAGCAGCGACGTCCTAGACATCCTGTGCCCATCCTCTTCTCCTCAACGATGCCCGGGCCGA TCAGTCT		
	ORF Start: ATG at 1	ORF Stop: end of sequence	

	SEQ ID NO: 82	414 aa	MW at 46487.9kD
NOV16a, CG54443-03 Protein Sequence	MGSTDSKLNFRKAVIQLTTKTQPV EATDDAFWDQFWADTATSVQDV FALVPAEIRAVREESPS NLATLCYKAVERLVQGAESGCHSEKEKQIVLNC SRLLTRVLPYIFEDPDWRGFFWSTVPGAGRG GQGEEDDEHARPLAESLLLAIDL LFCPDFTVQSHRRSTVDS AEDVHSLDSCEYIWEAGVGFH SPQPNYIHDNRMELLLKLLTFCSEMYLPPAPESGSTNPWVQFFCSTENRHALPLFTSLNLT CAYDPVGYGIPYNHLFSDYREPLVEEAQVLIVTL DHDSSASSASPTVDGTTGTAMDADPPG PENL FVNYLSRIHREDFQFILKGIARLLSNPL LQTYLPNSTKKIQFHELLVLFWKLCDFNKK FLFFVLKSSDVL DILVPI LFFLNDARADQS		

10

	SEQ ID NO: 83	1912 bp	
NOV16b, CG54443-07 DNA Sequence	CGAGGGCCGGGGGCGGGGCGCGCGCTTGTCTCTGCGAGAGCCGCGGGGCGCGGAGCTGGA GCCGGAGCTGAAGCCGGAGCCGGGTTGGAGTCTTGGGCGGGGCGGGGCGGAGCGGGCTCCAG AGACATGGGGTCCGACCGACTCCAAGCTGAACCTCCGGAAGGCGGTGATCCAGCTCACCACCAAG ACGCAGCCCGTGAAGCCACCGATGATGCCCTATGACCCTGTGGGTACGGGATCCCTACAACC ACCTGCTCTTCTCTGACACCGGGGAACCCCTGGTGGAGGAGGCTGCCAGGTGCTATTGTAC TTTGACCAAGACAGTGCCAGCAGTGCCAGCCCACTGTGGACGGCACCACCCTGGCACCAGCC ATGGATGATGCCGATCCTCCAGGCCCTGAGAACCTGTTGTGTAACCTGTGCCCATCCATC GTGAGGAGGACTTCCAGTTCATCCTCAAGGGTATAGCCCGGCTGCTGTCCAACCCCTGCTCCA GACCTACCTGCCCTAACTCCAAGAGATCCAGTTCACACGAGAGCTGTAGTTCTCTTCTGG AAGCTCTGCGACTTCAACAAGAAATCTCTCTTCTGCTGTAAGAGCAGCGACGTCTAGACA TCCTTGTCCCATCCTCTTCTTCTCAACGATGCCCGGGCGGATCAGTCTCGGGTGGGCTGAT GCACATTGGTGTCTTATCTTGTCTTCTGAGCGGGGAGCGGAACCTCGGGTGGGCTGAAC AAACCTACTCAATCCGCGTCCCATGGACATCCAGTCTTCAAGGAGCCACGCGGACCTGC TCATTGTGGTGTTCACAAGATCATCACCAGCGGGCACCAGCGGTTCAGCCCTCTTCGACTG CTGCTCACCATCGTGGTCAACGTGTCCCTACCTCAAGAGCCGTGTCATGGTGACCGCAAC AACCTGGTCTTCTTCTTCTTGGAGGTCTTCAACAACATCATCCAGTACAGTTTGTATGGCA CTCAACCTGGTCTACGCCATCATCCGCAAGCGCAGCATCTTCCACGAGCTGGCCAACTGCCC ACGACCCGCCCCACATTCAAGGCCCTGCAGCGCGCGCGGAGACCTGAGCCCTTGTCTC GCACCGGCTCCAGGAGGGCACTCCATGGAGGGCTCCCGCCCGCTGCCCTGCAGAGCCAGG CACCTCAAGACCACTGCTGGTGGCTACTCCAGGCATTGACAAGCTGACCGAGAAGTCCAGGTG TCAGAGGATGGCACCTTGCCTGCTCCGAACTGAGCCCGCAGCAGAGCTTGGAGGATGGCAGCC CGGCTAAGGGGAGCCAGCCAGGCATGGAGGGAGCAGCGCGACCATCCACCTCATCAGCCAG TGGCAGTGGAGCCCAACGCGCAGAGTGGGTCTCTCTGGAAGTCCGAAGCTGCCGTGCAGACC ATCATGAGGCTGCTGAGGTGCTGGTTCGCGAGGTGGAGAAGATCTGCATCGACAAGGGCTGA CGGATGAGTCTGAGATCCTGCGGTTCTGCGAGCATGGCACCTGGTGGGGCTGCTGCCCGTGC CCACCCATCCTCATCCGAAGTACAGGCCAACTCGGGCACTGCCATGTGGTTCGCACTAC ATGTGGGGCTCATCTATCTGAGGAATGTGGACCCCTGCTGTTGACGACCGGAGCTGAAGC TGTGTGAGATACAGCGGTGTGAGGATGAAGCCGACGAGGGGCTCAGTCTAGGGGAAGGAGGG CCTTGGTCCCTGAGGCTTCCCCATCCACATTCTGAGCTTTAAATTACCACGATC		
	ORF Start: ATG at 133		ORF Stop: TGA at 1813

	SEQ ID NO: 84	560 aa	MW at 63082.9kD
NOV16b, CG54443-07 Protein Sequence	MGSTSKLNFRAVIQLTKTQPV EATDDAYDPVGYGIPYNHLLFSDTGEPLVEEAAQVLIVTL DHDSASSASPTVDGTTTGTAMDDADPPGPENL FVNYSRIHREEDFQFILKGIARLLSNPLLOQ YLPNSTKIKIQHQLLVLEWKLCDENKFLFFVLKSSDVL DILVPIILFLINDARADQSRVGLMH IGVFILLLLSGERNFVRLNKPYSIRVPM DIPVFTGTHADLLIVVFHKIITSGHORLQPLFDCL LTIVNVNVPYLSLSMVTANKLLHLLEAFSTTWFLFSAAQNHHLVFFLELVFNNIIQYQFDGNS NLVYAIIRKRSIFHQLANLPTDPTIHKALQRRRTPEPLSR TGSQEGTSMEGSRPAAPAEFGT LKTSLVATPGIDKLTEKSVSEDTLRSLPEPEPQQLSDGSPAKGEPQAWREORRPSTSSASG QWSPTPEWVLSWKSPLPLQIMRLQVLVPQVEKICIDKGLTDESEILRFLQHGLVGLLPVPH PILIRKYQANS GTAMWFR T YMWGV IYLRNVDPVWYD TDVKLF EIQRV		

5

	SEQ ID NO: 85	3146 bp	
NOV16c, CG54443-01 DNA Sequence	ATGGGGTCGACCGACTCCAAGCTGAACCTCCGGAAGGCGGTGATCCAGCTCACCACCAAGACGC AGCCCGTGAAGCAACCGATGATGCCCTTTGGGACCACTTCTGGGCAGACACAGCCACCTCGGT GCAGGATGTGTTTGCACTGTGTCGCGCAGCAGAGATCCGGGCGGTGCGGGAAGAGTCAACCTCC AACTTGGCCACCTGTGCTACAAGGCCGTTGAGAAGCTGGTGACGGGAGCTGAGAGTGGCTGCC ACTCGGAGAAGGAGAAGCAGATCGTCTGAATGCGAGCGGCTGCTCACCCTGCTGCTGCCCTA CATCTTTGAGGACCCGACTGGAGGGGCTTCTTCTGGTCCAAGTGCACAGCAGGAGGAAGAG GATGATGAGCATGCCAGGCCCTGGCCGAGTCCCTGCTCTGGCATTGTGACCTGCTCTTCT GCCCGGACTTCACGGTTCAGAGCCACCGAGGAGCACTGTGACTCGGCAGAGGACGTCCACTC CCTGGACAGCTGTGAATACATCTGGGAGGCTGGTGTGGGCTTCGCTCACTCCCCCAGCCTAAC TACATCCACGATATGAACCGGATGGAGCTGCTGAAACTGCTGCTGACATGCTTCTCCGAGGCCA TGTACCTGCCCCAGCTCCGGAAGTGGCAGCACCAACCATGGGTTCAGTTCTTTTGTTCAC GGAGAACAGACATGCCCTGCCCTCTTCACTCCCTCTCAACACCGTGTGTGCCTATGACCT GTGGGCTACGGGATCCCTACAACCACTGCTCTTCTGACTACCGGGAACCCCTGGTGGAGG AGGCTGCCAGGTGCTCATTGTCACTTGGACCACGACAGTGCAGCATGCTTCTCCGAGGCCA GGACGGCACCAACCACTGGCACCGCCATGGATGATGCCGATGACTTCCAGTTCATCCTCAAGGGT ATAGCCCGGCTGCTGTCCAACCCCTGCTCCAGACCTACCTGCCTAACTCCACCAAGAGATCC		

	SEQ ID NO: 86	764 aa	MW at 86166.6kD
NOV16c, CG54443-01 Protein Sequence	MGS TDSK LNF RKA V I Q L T T K T Q P V E A T D D A F W D Q F W A D T A T S V Q D F V A L V P A A E I R A V R R E S P S N L A T L C Y K A V E K L V Q G A E S G C H S E K E Q I V L N C S R L L T R V L P Y I F E D P D W R G F F W S T V P Q Q G E E D D E H A R P L A E S L L L A I A D L L F C P D F T V Q S H R R S T V D S A E D V H S L D S C E Y I W E A G V G F A H S P Q P N Y I H D M N R M E L L K L L L T C F S E A M Y L P P A P E S G S T N P W V Q F F C S T E N R H A L P L F T S L I N T V C A Y D P V G Y G I P Y N H L L Q D Y R E P L V E E A A Q V I L V T L D H D S A S S A S P T V D G T T T G F A M D D A D D Q F T L K G I A R L L S N P L L Q T Y L P N S T K K I Q F H Q E L L V L F W K L C D F N K K F L F F V L K S S D V L D I L V P I L F F L N D A R A D Q S R V G L M H I G V F I L L L S G E R N F G V R L N K P Y S I R V P M D I P V P T G T H A D L L I V V F H K I I T S G H Q R L Q P L F D C L L T I V V N V S P Y L K S L S M V T A N K L L H L L E A F S T T W F L G S A A Q N H L V F F L L S V E N N I I Q Y Q F D G N S N L Y A I I R K R S I F H Q L A N L P T D P P T I H K A L Q R R R R T P E P L S R T G S Q E G T S M F G S R P A P A P E A P G T L S L V A T P G I D K L T E K S Q V S E D G T L R S L E P P Q S L E D G S P A K G E P S Q A W R E Q R R P S T P A E S G Q S W S P T P E W L V S W K S K L P L Q T I M R L L Q V L V P Q E K I C I D K G L T D E S E I L R F L Q H G T L V G L L P V P H P I L I R K Y Q A N S G T A M W F R T Y M G V I Y L R N V D P P V W Y D T D V K L F E I Q R V		

5

	SEQ ID NO: 87	3314 bp	
NOV16d, CG54443-02 DNA Sequence	GGCAGAGCCCGCGGGGGCCGCGGAGCTGGAGCCGGAGCTGAAGCCGGAGCCGGGTTGGAGTCTGG CGCGGGGGCCGGGCCGCGAGCGGGCTCCAGAGACATGGGGTGCACCGACTCCAAGTGAACCTCCG GAAGGCGGTGATCCAGCTCACCACCAAGACGCGCCCGTGGAAAGCCAGATGATGCTTCTTTGG GACCAGTTCTGGGCAGACACAGCCACCTCGGTGCAGGATGTGTTTGCACTGGTGCCGGCAGCAG AGATCCGGGGCCGTGCGGGAAGAGTCAACCTCCAACCTTGGCCACCCCTGTGCTACAGGCCGCTTGA GAAGCTGGTGCGAGGAGCTGAGAGTGGCTGCCACTCGGAGAAGGAGAAGCAGATCGTCTTGAAC TGCAGCCGGCTGCTCACCCGCGTGTGCCCTACATCTTTTGAGGAAACCCGACTGAGGGGGCTTCT TCTGGTCCACAGTGCCCGGGGCAGGGCGAGGAGGCGAGGAGAAGAGGATGATGAGCATGCCAG GCCCTTGGCCAGTCCCTGCTCCTGGCCATTGTGACACTGTCTTCTGCCCCGGACTTACGGTT CAGAGCCACCGGAGGAGCACTGTGGACTCGGCAGAGCAGTCACTGCTCCCTGGACAGCTGTGAAT ACATCTGGGAGGCTGGTGTGGGCTTCGCTCACTCCCCCAGCCTAACTACATCCACGATATGA		

	<p>CCGGATGGAGCTGCTGAAACTGCTGCTGACATGCTTCTCCGAGGCCATGTACCTGCCCCAGCT CCGGAAGTGGCAGCAACCAACCCATGGGTTCACTTCTTTTGTTCACGGAGAACAGACATGCC TGCCCTCTTCACTCCCTCCTCAACACCGTGTGTGCTATGACCTGTGGGCTACGGGATCCC CTACAACCACTGCTCTTCTCTGACACCGGGGAACCCCTGGTGGAGGAGGCTGCCAGGTGCTC ATTGTCACTTTGGACCAAGACAGTGCAGTGCAGCAGTGCACCGGCTGTGGACGGCACCACTG GCACCGCCATGGATGATGCGGATCCTCCAGGCCCTGAGAACCTGTTTGTGAACTACCTGTCCC CATCCATCGTGAGGAGGACTTCCAGTTTCTCTCAAGGGTATAGCCCCGCTGTGTCCAAACCC CTGCTCCAGACCTACCTGCTTAACTCCACCAAGAAGATCCAGTTCACCAAGAGCTGTAGTTT TCTTCTGGAAGCTCTGCGCACTTCAACAAGAAATTCCTTCTTCTGCTGCTGAAGAGCAGCG CCTAGACATCCTTGTCCCCATCCTCTTCTTCTCTCAACGATGCCCGGGCCGATCAGTCTCGGGT GGCTGATGCACATTGGTGTCTTCTCTTCTGCTGCTTCTGAGCGGGGAGCGGAACCTCGGGGTG GGCTGAACAAACCTACTCAATCCGCGTGCCCATGGACATCCCACTTCAACAGGACCCACGC CGACCTGCTCATTTGGTGTGTTCCACAAGATCATCACCGGGCAGCAGCGGTTCAGCCCTC TTGACTGCTGCTCATCACCATCGTGGTCAACGTGTCCCCCTACCTCAAGAGCCTGTCCATGGTGA CCGCCAACAAAGTTGCTGCACTGCTGGAGGCTTCTCCACCACTGGTCTCTTCTCTGCCG CCAGAACCACCACTGGTCTTCTTCTCTGAGGCTTCAACAACATCATCCAGTACCACTTT GATGGCAACTCCAACCTGGTCTACGCCATCATCCGCAAGCGCAGCATCTTCCACAGCTGGCCA ACCTGCCCAACCGACCCGCCACCAATTCAACAGGCCCTGCAGCGGCGCCGCGGACACCTGAGCC CTGTCTCGCACCGGCTCCAGGAGGGCACCTCCATGGAGGGTCCCCGCCCGCTGCCCTGCA GAGCCAGGACCCCTCAAGACCAAGTCTGGTGGCTACTCCAGGCAATTGACAAGCTGACCGAGAAGT CCCAGGTGTGAGGATGGCAGCTTGGCGGTCCCTGGAACTGAGCCCCAGCAGAGCTTGGAGGA TGGCAGCCCGGCTAAGGGGGAGCCAGCCAGGATGGAGGGAGCAGCGGCGACCATCCACCTCA TCAGCCAGTGGGCACTGGAGCCCAACGCCAGAGTGGGTCTCTCTGGAAGTGAAGCTGCCG TGACAGCATCATGAGGCTGCTGCAAGTGTGTTCCGAGGTGGAGAAGATCTGCATCGACAA GGGCTGACGGATGAGTCTGAGATCCTGCGGTTCTGCGAGCATGGCACCTGGTGGGGCTGCTG CCCGTGCCCCACCCATCCTCATCCGCAAGTACAGGCCAATCGGGCACTGCCATGTGGTTCC GCACCTACATGTGGGGCGTCTATCTGAGGAATGTGGACCCCTGTCTGGTACGACACCGA CGTGAAGCTGTTTGAGATA CAGCGGGTGTGAGGATGAAGCGGACGAGGGGCTCAGTCTAGGGA AGGCAGGGCTTGGTCCCTGAGGCTTCCCCATCCACCATCTGAGCTTTAAATTACCAAGATC AGGGCTTGAACAGGCAGAGTGGCCCTGAGTGTCTAGCCCTAGAGACCCCTGTGGCCAGGACAA TGTGAATGGCTCAGATCCCCCTCAACCCCTAGGCTGGACTCAGACGAGCCCATCTCTGGGCG TATGCCCCCAGCAGACCACTGCCCCCAACACTCGGACTCCCTCTTTAAGACCTGGCTCAGTG CTGGCCCTCAGTGCCCAACCACTCCTGTGCTACCCAGCCCCAGAGGCAGAACCCAAATGGGT CACTGTGCCCTAAGGGGTTTGACAGGGAAACACGGGCTGTCCCTTGGAGTGCCTGGACAGGGT AAGGGGGTGTCTCCAGCTCCTTAACCCAAAGCCAGCTGTTTCCAGGCTCAGGGGAGAAAGGTGT GGCCAGGCTGCTCCTCGAGGAGCTGGGAGCTGGCCGACTGCAAAAGCCAGACTGGGGCACCTC CCGTATCCTTGGGGCATGGTGTGGGGTGGTGGGGTCTCCTGCTATATTCTCTGGATCCATGG AAATAGCCTGGCTCCCTCTTACCAGTAATGAGGGGACGGGAAGGAACTGGGAGGCAGCCGT TAGTCTCCCTGCCCTGCCACTGCTGGATGGGGCGATGCCACCCCTCATCTTCAACCCAGCT CTGGCCTCTGGGTCCCAACCCAGCCCCCGTGTGAGAACATCTTTGCTCTGTACAATCGG CTCTTTACAATAAAACCTCTGCTCCAAAAA</p>	
	ORF Start: ATG at 97	ORF Stop: TGA at 2461

	SEQ ID NO: 88	788 aa	MW at 88582.2kD
NOV16d, CG54443-02 Protein Sequence	<p>MGSTDSKLNFRKAVIQLTTKTPVEATDDAFWDQFWADTATSVQDVFAVPAAEIRAVREESPS  NLATLCYKAVEKLVQGAESGCHSEKEQIVLNCRRLLTRVLPYIFEDPDWRGFFWSTVPAGRG  GQGEEDDEHARPLAESLLLAIDLLFCPDTVQSHRRSTVDSAEVDHSLDSCEYIWEAGVGFAH  SPQPNYIHDNMNRMELLKLLTFCSEAMYLPPAPESGSTNFWVQFPCSTENRHALPLFTSLNNTV  CAYDPVGYGIPYNHLLFSDTGEPLVEEAAQVLIVTLDDHDSASSASPTVDGTTTGTAMDDADPPG  PENLNVNYSRIHREDFQFILKGIARLLSNPLLOTYLPNSTKKIQFHQELLVLFWKLCDFNKK  FLFFVLKSSDVLIDILVPIILFNDARADQSRVGLMHIGVFILLLSGERNFVRLNKPYSIRVP  MDIPVFTGTHADLLIVVFHKTITSGHQRLOPLFDCLLTIVNVSPYLKSLSMVTANKLLHLLLEA  FSTTWFLFSAQNHLVFFLLEVFNNIIQYQFDGNSNLVYAIIRKRSIFHQLANLPTDPPTIHK  ALQRRRTPEPLSRGSGEQTSMEGSRPAAPAEPTLKTSLVATPGIDKLTESQVSEDGTLRS  LEPEPQQLSDGSPARGESQAWREQRRPSTSSAGQWSPTEWVLSWKSPLQTIMRLQLVL  VPQVEKICIDKGLTDESEILRFLQHGTLVGLLPVPHPIILIRKYQANSQTAMNFTYMWGVITYLR  NVDPPVWYDTDKVLFETQRV</p>		

5

	SEQ ID NO: 89	1242 bp
NOV16e, CG54443-04 DNA	<p>ATGGGGTTCGACCGACTCCAAGCTGAACCTCCGGAAGGCGGTGATCCAGCTCACCACCAAGACGCA  GCCCGTGAAGCCACCGATGATGCTTTTGGGACCACTCTGGGCAGACACGCCACCTCGGTGC  AGGATGTGTTTGCACTGGTGCCGCGCAGCAGAGATCCGGGCCGTGCGGGAAGAGTCAACCTCCAAC</p>	

Sequence	TTGGCCACCTGTGTCTACAAGGCCGTTGAGAGGCTGGTGACGGGAGCTGAGAGTGGCTGCCACTC GGAGAAGGAGAAGCAGATCGTCTGAACTGCAGCCGGCTGCTACCCGCGTGCTGCCCTACATCT TTGAGGACCCCGACTGGAGGGGCTTCTTCTGGTCCACAGTGCCCCGGGAGCGGGCGAGGAGGGCAG GGAGAAGAGGATGATGAGCATGCCAGGCCCTGGCCGAGTCCCTGCTCCTGGCCATTGTGTGACCT GCTCTTCTGCCCGGACTTCACGGTTTACAGAGCCACCGGAGGAGCACTGTGGACTCGGCAGAGGACG TCCACTCCCTGGACAGCTGTGAATACATCTGGGAGGCTGGTGTGGGCTTCGCTCACTCCCCCAG CCTAACTACATCCAGATATGAACCGGATGGAGCTGCTGAAACTGCTGCTGACATGCTTCTCCGA GGCCATGTACCTGCCCCCAGCTCCGAAAGTGGCAGCACCAACCATGGGTTCACTTCTTTGTT CCACGGAGAACAGACATGCCCTGCCCTCTTCACTCCCTCCTCAACCCGTGTGTGCCTATGAC CCTGTGGGCTACGGGATCCCCTACAACCACCTGCTCTTCTCTGACTACCGGGAAACCCCTGGTGA GGAGGCTGCCAGGTGCTATTGTCACTTTGGACACGACAGTGCAGCAGTGCACGCCCTCACTG TGGACGGCACCACCACTGGCACCGCCATGGATGATGCCGATCCTCCAGGCCCTGAGAACCTGTTT GTGAACCTACCTGTCCCGCATCCATCGTGAGGAGGACTTCCAGTTCATCCTCAAGGGTATAGCCCG GCTGCTGTCCAACCCCTGCTCCAGACCTACCTGCCTAACTCCACCAAGAGATCCAGTTCCACC AGGAGCTGCTAGTTCTTCTGGAAGCTCTGCGACTTCAACAAGAAATTCCTCTTCTCGTGTG AAGAGCAGCGACGTCTAGACATCCTTGTCCCCATCCTTCTTCTCTCAACGATGCCCGGGCCGA TCAGTCT	
	ORF Start: ATG at 1	ORF Stop: end of sequence

	SEQ ID NO: 90	414 aa	MW at 46487.9kD
NOV16e, CG54443-04 Protein Sequence	MGSTD SKLNFRKAVIQLTTKTPVEATDDAFWDQFWADTATSVQDV FALVPAAEIRAVREESPS NLATLCYKAVERLVQGAESGCHSEKEKQIVLNC SRLLTRVLPYIFEDPDWRGFFWSTVPGAGRG GQGEEDDEHARPLAESLLAIADLLFCPDFTVQSHRRSTVDSAEVDHSLDSEYIWEAGVGFAH SPQPNYIHD MNRMELLKLLLTCFSEAMYLPPAPESGSTNPWVQFFCSTENRHALPLFTSLNLT CAYDPVGYGIPYNHLLFSYREPLVEEAAQVLI VTL DHDSSASSASPTVDGTTTGTAMDADPPG PENLFVNYLSRIHREEDFQFILKGIARLLSNPLLQTYLPNSTKKIQFHQELLVLFWKLCDFNKK FLFFVLKSSDVL DILVPIFLFLNDRADQS		

5

	SEQ ID NO: 91	1242 bp
NOV16f, CG54443-05 DNA Sequence	ATGGGGTCGACCGACTCCAAGCTGAACTTCCGGAAGGCGGTGATCCAGCTCACCACCAAGACGCA GCCC GTGGAAGCCACCGATGATGCCCTTTGGGACCACTTCTGGGCAGACACAGCCACCTCGGTGC AGGATGTGTTTGCACTGGTGCCGCGCAGCAGAGATCCGGGCCGTGCGGGAAGAGTCACCCCTCCAAC TTGGCCACCCTGTGCTACAAGGCCGTTGAGAGGCTGGTGACAGGAGCTGAGAGTGGCTGCCACTC GGAGAAGGAGAAGCAGATCGTCTGAACTGCAGCCGGCTGCTACCCGCGTGTGCCCTACATCT TTGAGGACCCCGACTGGAGGGGCTTCTTCTGGTCCACAGTGCCCGGGGAGGGCGAGGAGGGCAG GGAGAAGAGGATGATGAGCATGCCAGGCCCTGGCCGAGTCCCTGCTCTGCGCCATTGCTGACCT GCTCTTCTGCCCGGACTTCACGGTTTACAGGCCACCGAGGAGCACTGTGGACTCGGCAGAGGACG TCCACTCCCTGGACAGCTGTGAATACATCTGGGAGGCTGGTGTGGGCTTCGCTCACTCCCCCAG CCTAACTACATCCAGATATGAACCGGATGGAGCTGCTGAACTGCTGTGACATGCTTCTCCGA GGCCATGTACCTGCCCCCAGCTCCGGAAGTGGCAGCACCAACCCATGGGTTCACTTCTTTGTT CCACGGAGAACAGACATGCCCTGCCCTCTTCACTCCCTCCTCAACACCGTGTGTGCCTATGAC CCTGTGGGCTACGGGATCCCCTACAACCACCTGCTTCTCTGACTACCGGGAACCCCTGGTGA GGAGGCTGCCAGGTGCTCATTTGTCACTTTGGACACGACAGTGCCAGCAGTGCCAGCCCCACTG TGGACGGCACCACCACTGGCACCGCCATGGATGATGCCGATCCTCCAGGCCCTGAGAACCCTGTTT GTGAACCTACCTGTCCCGCATCCATCGTGAGGAGGACTTCCAGTTTCCTCAAGGGTATAGCCCG GCTGCTGTCCAACCCCTGCTCCAGACCTACCTGCCTAACTCCACCAAGAAGATCCAGTTCCACC AGGAGCTGCTAGTTCTTCTTCTGGAAGCTCTGCGACTTCAACAAGAAATTCCTCTTCTCTGCTG AAGAGCAGCGACGTCCTAGACATCCTTGTCCCCATCCTCTTCTCTCAACGATGCCCGGGCCGA TCAGTCT	
	ORF Start: ATG at 1	ORF Stop: end of sequence

	SEQ ID NO: 92	414 aa	MW at 46487.9kD
NOV16f, CG54443-05 Protein Sequence	MGSTD SKLNFRKAVIQLTTKTPVEATDDAFWDQFWADTATSVQDV FALVPAAEIRAVREESPS NLATLCYKAVERLVQGAESGCHSEKEKQIVLNC SRLLTRVLPYIFEDPDWRGFFWSTVPGAGRG GQGEEDDEHARPLAESLLAIADLLFCPDFTVQSHRRSTVDSAEVDHSLDSEYIWEAGVGFAH SPQPNYIHD MNRMELLKLLLTCFSEAMYLPPAPESGSTNPWVQFFCSTENRHALPLFTSLNLT CAYDPVGYGIPYNHLLFSYREPLVEEAAQVLI VTL DHDSSASSASPTVDGTTTGTAMDADPPG PENLFVNYLSRIHREEDFQFILKGIARLLSNPLLQTYLPNSTKKIQFHQELLVLFWKLCDFNKK		

FLFFVLKSSDVL DILVPI LFFLNDARADQS
----------------------------------

	SEQ ID NO: 93	1242 bp
NOV16g, CG54443-06 DNA Sequence	ATGGGGTCGACCGACTCCAAGCTGAACCTCCGGAAGGCGGTGATCCAGCTCACCACCAAGACGCA GCCCGTGGGAAGCCACCGATGATGCCTTTTGGGACCAGTTCTGGGCAGACACAGCCACCTCGGTGC AGGATGTGTTTGCACTGGTGCCGGCAGCAGAGATCCGGGCGGTGCGGGAAGAGTCAACCTCCAAC TTGGCCACCCTGTGCTACAAGSCGTTGAGAGGCTGGTGACGGGAGCTGAGAGTGGCTGCCACTC GGAGAAGGAGAAGCAGATCGTCTGAACTGCAGCCGGCTGCTCACC CGCGTGTGCCCTACATCT TTGAGGACCCCGACTGGAGGGGCTTCTTCTGGTCCACAGTGCCCGGGGAGGGCGAGGGGCGAG GGAGAAGAGGATGATGAGCATGCCAGGCCCTGGCCGAGTCCCTGCTCCTGGCCATTGCTGACCT GCTCTTCTGCCCGGACTTCACGGTTCAGAGCCACCGGAGGAGCACTGTGGACTCGGCAGAGGACG TCCACTCCCTGGACAGCTGTGAATACATCTGGGAGGCTGGTGTGGGCTTCGCTCACTCCCCCAG CCTAAC TACATCCAGATATGAACCGGATGGAGCTGCTGAACTGCTGCTGACATGCTTCTCCGA GGCCATGTACCTGCCCCAGCTCCGGAAAGTGGCAGCACCAACCCATGGGTTCAAGTCTTTTGT CCACGGAGAACAGACATGCCCTGCCCTCTTCACTCCCTCCTCAACACCGTGTGTGCTATGAC CCTGTGGGCTACGGGATCCCTACACCACTGCTCTTCTGACTACCGGAAACCCCTGGTGGG GGAGGCTGCCCCAGGTGCTCATTGTCACTTTGGACCACGACAGTGCCAGTGCAGGCCCACTG TGGACGGCACCACCACTGGCACCAGCATGGATGATGCCGATCCTCCAGGCCCTGAGAACCTGTTT GTGAAC TACCTGTCCCGCATCCATCGTGAGGAGGACTTCCAGTTCATCCTCAAGGGTATAGCCCG GCTGCTGTCAACCCCTGCTCCAGACCTACCTGCCCTAACTCCACCAAGAAGATCCAGTCCACC AGGAGCTGCTAGTTCTTCTTCTGGAAGCTCTGCGACTTCAACAAGAAATCTCTTCTGCTGCTG AAGAGCAGCGACGTCTAGACATCCTTGTCCCATCTTCTTCTCTCAACGATGCCCGGGCCGA TCAGTCT	
	ORF Start: ATG at 1	ORF Stop: end of sequence

5

	SEQ ID NO: 94	414 aa	MW at 46487.9kD
NOV16g, CG54443-06 Protein Sequence	MGSTDSKLNFRKAVIQLTTKTQPV EATDDAFWDQFWADTATSVQDVFALVPAEIRAVREESPS NLATLCYKAVERLVQGAESGCHSEKEKQIVLNC SRLLTRVLPYIFEDPDWRGFFWSTVPGAGRG GQGEEDDEHARPLAESLLAIADLLFCPDFTVQSHRRSTVDSAEVDVHSLDSEYIWEAGVGFH SPQPNYIHD MNRMELLKLLTFCFSEAMYLPPAPESGSTNPNWVQFPCSTENRHALPLFTSLNTV CAYDPVGYGIPYNHLLFS DYREPLVEEAAQVLI VTL DHD SASSASPTVDGTTGTAMDDADPPG PENL FVN YLSRIHREEDFQFILKGIARLLSNPILQTYLPNSTKKIQFHQELLVLFWKLCDFNKK FLFFVLKSSDVL DILVPI LFFLNDARADQS		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 16B.

10

Table 16B. Comparison of NOV16a against NOV16b through NOV16g.		
Protein Sequence	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV16b	258..414 30..186	155/157 (98%) 155/157 (98%)
NOV16c	1..414 1..390	388/414 (93%) 389/414 (93%)
NOV16d	1..414 1..414	411/414 (99%) 412/414 (99%)
NOV16e	1..414 1..414	414/414 (100%) 414/414 (100%)
NOV16f	1..414 1..414	414/414 (100%) 414/414 (100%)
NOV16g	1..414 1..414	414/414 (100%) 414/414 (100%)

Further analysis of the NOV16a protein yielded the following properties shown in Table 16C.

5

Table 16C. Protein Sequence Properties NOV16a	
SignalP analysis:	No Known Signal Sequence Indicated
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 11; pos.chg 2; neg.chg 1  H-region: length 0; peak value -0.21  PSG score: -4.61</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): -5.12  possible cleavage site: between 48 and 49</p> <p>&gt;&gt;&gt; Seems to have no N-terminal signal peptide</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 1  Tentative number of TMS(s) for the threshold 0.5:  2  Number of TMS(s) for threshold 0.5: 0  PERIPHERAL Likelihood = 3.13 (at 93)  ALOM score: -1.28 (number of TMss: 0)</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 1 Hyd Moment (75): 2.33  Hyd Moment (95): 1.80 G content: 1  D/E content: 2 S/T content: 6  Score: -5.83</p>

	<p>Gavel: indication of cleavage sites for mitochondrial preseq R-2 motif at 21 FRK AV</p> <p>NUCDISC: discrimination of nuclear localization signals pat4: none pat7: none bipartite: none content of basic residues: 7.7% NLS Score: -0.47</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals: none</p> <p>SKL: peroxisomal targeting signal in the C-terminus: none</p> <p>PTS2: 2nd peroxisomal targeting signal: none</p> <p>VAC: possible vacuolar targeting motif: none</p> <p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif: type 1: none type 2: none</p> <p>NMYR: N-myristoylation pattern : MGSTDSK</p> <p>Prenylation motif: none</p> <p>memYQRL: transport motif from cell surface to Golgi: none</p> <p>Tyrosines in the tail: none</p> <p>Dileucine motif in the tail: none</p> <p>checking 63 PROSITE DNA binding motifs: none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p> <p>checking 33 PROSITE prokaryotic DNA binding motifs: none</p> <p>NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination Indication: cytoplasmic Reliability: 94.1</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions total: 0 residues</p> <p>-----</p> <p>Final Results (k = 9/23):</p> <p>43.5 %: cytoplasmic 30.4 %: mitochondrial</p>
--	---

	21.7 %: nuclear 4.3 %: peroxisomal  >> indication for CG54443-03 is cyt (k=23)
--	---

A search of the NOV16a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 16D.

5

Table 16D. Geneseq Results for NOV16a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG70273	Human CG8841-like protein #2 - Homo sapiens, 788 aa. [WO200255702-A2, 18-JUL-2002]	1..414 1..414	411/414 (99%) 412/414 (99%)	0.0
AAM79253	Human protein SEQ ID NO 1915 - Homo sapiens, 787 aa. [WO200157190-A2, 09-AUG-2001]	1..414 1..413	412/414 (99%) 413/414 (99%)	0.0
ABG70272	Human CG8841-like protein #1 - Homo sapiens, 764 aa. [WO200255702-A2, 18-JUL-2002]	1..414 1..390	388/414 (93%) 389/414 (93%)	0.0
ABB12112	Human secreted protein homologue, SEQ ID NO:2482 - Homo sapiens, 284 aa. [WO200157188-A2, 09-AUG-2001]	1..271 12..283	259/272 (95%) 261/272 (95%)	e-151
ABB64025	Drosophila melanogaster polypeptide SEQ ID NO 18867 - Drosophila melanogaster, 837 aa. [WO200171042-A2, 27-SEP-2001]	1..414 1..398	252/414 (60%) 310/414 (74%)	e-146

In a BLAST search of public sequence databases, the NOV16a protein was found to have homology to the proteins shown in the BLASTP data in Table 16E.

10

Table 16E. Public BLASTP Results for NOV16a				
Protein Accession Number	Protein/Organism/Length	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAH35372	Hypothetical protein - Homo sapiens (Human), 788 aa.	1..414 1..414	413/414 (99%) 414/414 (99%)	0.0
Q8TE83	Hypothetical protein FLJ23821 - Homo sapiens (Human), 625 aa.	1..414 1..414	413/414 (99%) 414/414 (99%)	0.0
Q8R1F6	Hypothetical 88.8 kDa protein - Mus musculus (Mouse), 788 aa.	1..414 1..414	400/414 (96%) 410/414 (98%)	0.0
Q9NT34	Hypothetical protein - Homo sapiens (Human), 380 aa (fragment).	1..364 1..363	354/364 (97%) 355/364 (97%)	0.0
Q9V695	CG8841 protein - Drosophila melanogaster (Fruit fly), 837 aa.	1..414 1..398	252/414 (60%) 310/414 (74%)	e-146

### Example 17.

- 5 The NOV17 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 17A.

Table 17A. NOV17 Sequence Analysis			
	SEQ ID NO: 95	752 bp	
NOV17a, CG58495-01 DNA Sequence	CTCTGGCCCTCACCCTCATCTTGATGGCAGCCTCTGGTGTGCGTGCAGAGTGAGGGACGTTTG TGTTGGAAGCCCTGGTATCCCCGGCACTCCTGGATCCACGGCCTGCCAGGCAGGGACGGGAGA GATGGTGTCAAAGGAGACCTTGGCCCTCCAGGCCCATGGGTCCGCTGGAGAAACACCATGTC CTCCTGGGAATAATGGGCTGCTGGAGCCCTGGTGTCCCTGGAGAGCGTGGAGAGAAGGGGGA GCCTGGCGAGAGAGGCCCTCCAGGCTTCCAGCTCATCTAGATGAGGAGCTCCAAGCCCACTC CAGCACTTCAGACATCAAATCCTGCAGACAAGGGGAGCCCTCAGTCTGCAGGGCTCCATAATGA CAGTAGGAGAGAAGGTCTTCTCCAGCAATGGGCAGTCCATCACTTTGATGCCATTTCAGGAGGC ATGTGCCAGAGCAGGCGGCCGATTGCTGTCCCAAGGAATCCAGAGGAAAATGAGGCCATTGCA AGCTTCGTGAAGAAGTACAACACATATGCCTATGTAGGCCTGACTGAGGGTCCAGCCCTGGAG ACTTCCGCTACTCAGATGGGACCCCTGTAAACTACACCAACTGGTACCGAGGGGAGCCTGCAGG TCGGGAAAAGAGAAGTGTGTGGAGATGTACACAGATGGGCAGTGAATGACAGGAATGCGCTG TACTCCCGACTGACCATCTGTGAGTTCTGAGAGGCATTTAGGCCATGG		
	ORF Start: at 3		ORF Stop: TGA at 732

	SEQ ID NO: 96	243 aa	MW at 25592.3kD
NOV17a, CG58495-01 Protein Sequence	LALTLLILMAASGAACEVRDVCVGSFPIPGTSGHGLPGRDRDGVKGDGPPGPMGPGETPCP PGNNGLPGAPGVPGERGEKGEPPERGPPLPAHLDEELQATLHDFRHQILQTRGALSQGSI MVTGEKVFSSNGQSITFDALQIEACARAGGRIAVPRNPEENEAIASFVKYNTYAVVGLTEGSPSGD FRYSDGTPVNYTNWYRGEPAGRGKEKCVEMYTDGQWDRNCLYSRLTICEF		

	SEQ ID NO: 97	681 bp	
NOV17b, CG58495-03 DNA Sequence	CCAAGCACCTGGAGGCTCTGTGTGTGGGTCGCTGATTTCTTGGAGCCTGAAAAGAAGGAGCAGC GACTGGACCCAGAGCCATGTGGCTGTGCCCTCTGGCCCTCACCTCATCTTGATGGCAGCCTCT GGTGCTGCGTGCGAAGTGAAGGAGCTCCAAGCCACACTCCACGACTTCAGACATCAAATCCTGC AGACAAGGGGAGCCCTCAGTCTGCAGGGCTCCATAATGACAGTAGGAGAGAAGGTCTTCTCTAG CAATGGGCAGTCCATCACTTTTGATGCCATTGAGGAGGATGTGCCAGAGCAGGCGGCGCATT GCTGTCCCAAGGAATCCAGAGGAAAATGAGGCCATTGCAAGCTTCGTGAAGAAGTACAACAT ATGCCTATGTAGGCTGACTGAGGGTCCAGCCCTGGAGACTTCCGCTACTCAGATGGGACCCC TGTAACTACACCACTGGTACCGAGGGGAGCCTGCAGGTCGGGGAAGAGAGAGTGTGTGGAG ATGTACACAGATGGGCAGTGAATGACAGGAATGCTGTACTCCCGACTGACCATCTGTGAGT TCTGAGAGGCATTAGGCCATGGGACAGGGAGGATCCTGTCTGGCCTTCAGTTTCCATCCCCAG GATCCACTTGGTCTGTGAGATGCTAGAACTCCCTTTCAACA		
	ORF Start: ATG at 81		ORF Stop: TGA at 579

5

	SEQ ID NO: 98	166 aa	MW at 18388.6kD
NOV17b, CG58495-03 Protein Sequence	MWLCPLALTLILMAASGAACEVKELQATLHDFRHQILQTRGALSLSQSIMTVGEKVFSSNGQSI TFDALQIEACARAGGRIAVPRNPEENEALASFVKYNTYAYVGLTEGSPGDFRYSDDGTPVNYTN WYRGEPAGRGKEKCVEMYTDGQWNRNCLYSRLTICEF		

	SEQ ID NO: 99	1161 bp	
NOV17c, CG58495-02 DNA Sequence	GGCTCTTTCTAGCTATAAACACTGCTTGCCGCGCTGCACTCCACCACGCTCCTCCAAGTCCCA GCGAACCCGCGTGCAACCTGTCCCGACTCTAGCCGCTCTTCAGCTCACGGATCAATTCCTCAAG TCGCTGGAGGCTCTGTGTGTGGGAGCAGCGACTGGACCCAGAGCCATGTGGCTGTGCCCTCTGG CCTCAACCTCATCTTGATGGCAGCCTCTGGTGCTGTGTGCGAAGTGAAGGACGTTGTGTGG AAGCCCTGGTATCCCCGGCACTCCTGGATCCCACGGCCTGCCAGGCAGGGACGGGAGAGATGGT GTCAAAGGAGACCCCTGGCCCTCCAGGCCCATGGGTCCACCTGGAGAAAATGCCATGTCTCTCT GAAATGATGGGCTGCCTGGAGCCCTGGTATCCCTGGAGAGTGTGGAGAGAAGGGGAGCCTGG CGAGAGGGGCCCTCCAGGGCTTCAGCTCATCTAGATGAGGAGCTCCAAGCCACTCCACGAC TTTAGACATCAAATCCTGCAGACAAGGGGAGCCCTCAGTCTGCAGGGCTCCATAATGACAGTAG GAGAGAAGGTCTTCTCCAGCAATGGGCAGTCCATCACTTTTGATGCCATTGAGGAGCATGTGC CAGAGCAGGCGGCCGATTGCTGTCCCAAGGAATCCAGAGGAAAATGAGGCCATTGCAAGCTTC GTGAAGAAGTACAACATATGCCTATGTAGGCCTGACTGAGGGTCCCAGCCCTGGAGACTTCC GCTACTCAGACGGGACCCCTGTAACTACACCACTGGTACCAGGGGAGCCCGCAGGTCCGGG AAAAGAGCAGTGTGTGGAGATGTACACAGATGGGCAGTGAATGACAGGAACTGCCCTGTACTCC CGACTGACCATCTGTGAGTTCTGAGAGGCATTAGGCCATGGGACAGGGAGGACGCTCTCTGGC CTCCATCCTGAGGCTCCACTTGGTCTGTGAGATGCTAGAACTCCCTTCAACAGAATTGATCCCT GCTGCCCCTGCTGGAGAGCTTCAAGGTGAGCTTCTGAGCGCTCTCTGAGGAGTACACTAAGA AGCTCAACACCCAGTGAGGCGCCCGCCGCGCCCTTCCCGTGCTCAGAATAAACGTTTCC AAAGTGGGA		
	ORF Start: ATG at 174		ORF Stop: TGA at 918

10

	SEQ ID NO: 100	248 aa	MW at 26228.2kD
NOV17c, CG58495-02 Protein Sequence	MWLCPLALNLILMAASGAVCEKDVCGVSPGIPGTPGSHGLPGRDGRDGVKDPGPPGPMGPPG EMPCPGNDGLPGAPGIPGECGEKGEPPGERGPPGLPAHLDEELQATLHDFRHQILQTRGALSQ GSIMTVGEKVFSSNGQSI TFDALQIEACARAGGRIAVPRNPEENEALASFVKYNTYAYVGLTEG PSPGDFRYSDDGTPVNYTNWYRGEPAGRGKEQCVEMYTDGQWNRNCLYSRLTICEF		

Sequence comparison of the above protein sequences yields the following sequence  
 15 relationships shown in Table 17B.

Table 17B. Comparison of NOV17a against NOV17b and NOV17c.		
Protein Sequence	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV17b	100..243 23..166	143/144 (99%) 144/144 (99%)
NOV17c	1..243 6..248	235/243 (96%) 239/243 (97%)

Further analysis of the NOV17a protein yielded the following properties shown in Table 17C.

5

Table 17C. Protein Sequence Properties NOV17a	
SignalP analysis:	Cleavage site between residues 16 and 17
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 0; pos.chg 0; neg.chg 0  H-region: length 15; peak value 10.71  PSG score: 6.31</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): 2.44  possible cleavage site: between 15 and 16</p> <p>&gt;&gt;&gt; Seems to have a cleavable signal peptide (1 to 15)</p> <p>ALOM: Klein et al's method for TM region allocation  Init position for calculation: 16  Tentative number of TMS(s) for the threshold 0.5:  0  number of TMS(s) .. fixed  PERIPHERAL Likelihood = 7.37 (at 113)  ALOM score: 7.37 (number of TMSs: 0)</p> <p>MTOP: Prediction of membrane topology (Hartmann et al.)  Center position for calculation: 7  Charge difference: -2.0 C(-1.0) - N( 1.0)  N &gt;= C: N-terminal side will be inside</p> <p>MITDISC: discrimination of mitochondrial targeting seq  R content: 0 Hyd Moment (75): 2.24  Hyd Moment (95): 0.60 G content: 1  D/E content: 1 S/T content: 2  Score: -6.05</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq  cleavage site motif not found</p> <p>NUCDISC: discrimination of nuclear localization signals  pat4: none</p>

	pat7: none bipartite: none content of basic residues: 9.1% NLS Score: -0.47  KDEL: ER retention motif in the C-terminus: none  ER Membrane Retention Signals: none  SKL: peroxisomal targeting signal in the C-terminus: none  PTS2: 2nd peroxisomal targeting signal: none  VAC: possible vacuolar targeting motif: none  RNA-binding motif: none  Actinin-type actin-binding motif: type 1: none type 2: none  NMYR: N-myristoylation pattern : none  Prenylation motif: none  memYQRL: transport motif from cell surface to Golgi: none  Tyrosines in the tail: none  Dileucine motif in the tail: none  checking 63 PROSITE DNA binding motifs: none  checking 71 PROSITE ribosomal protein motifs: none  checking 33 PROSITE prokaryotic DNA binding motifs: none  NNCN: Reinhardt's method for Cytoplasmic/Nuclear discrimination Indication: nuclear Reliability: 55.5  COIL: Lupas's algorithm to detect coiled-coil regions total: 0 residues  -----  Final Results (k = 9/23):  66.7 %: extracellular, including cell wall 22.2 %: nuclear 11.1 %: mitochondrial  >> indication for CG58495-01 is exc (k=9)
--	---

A search of the NOV17a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 17D.

<b>Table 17D. Geneseq Results for NOV17a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV17a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAU76468	Human lung surfactant protein A - Homo sapiens, 248 aa. [WO200206301-A2, 24-JAN-2002]	1..243 6..248	241/243 (99%) 243/243 (99%)	e-147
AAU77989	Human SP-A amino acid sequence - Homo sapiens, 248 aa. [WO200011161-A1, 02-MAR-2000]	1..243 6..248	241/243 (99%) 243/243 (99%)	e-147
AAP70662	35kd pulmonary surfactant protein - Homo sapiens, 248 aa. [WO8702037-A, 09-APR-1987]	1..243 6..248	240/243 (98%) 242/243 (98%)	e-146
AAR05091	Vector PSP 35K-1A-10 gene product encoding pulmonary surfactant protein - Homo sapiens, 248 aa. [US4882422-A, 21-NOV-1989]	1..243 6..248	239/243 (98%) 242/243 (99%)	e-146
AAB58135	Lung cancer associated polypeptide sequence SEQ ID 473 - Homo sapiens, 259 aa. [WO200055180-A2, 21-SEP-2000]	1..243 17..259	238/243 (97%) 239/243 (97%)	e-145

5

In a BLAST search of public sequence databases, the NOV17a protein was found to have homology to the proteins shown in the BLASTP data in Table 17E.

<b>Table 17E. Public BLASTP Results for NOV17a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV17a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
LNHUP1	pulmonary surfactant protein A precursor (clone 1A) - human, 248 aa.	1..243 6..248	240/243 (98%) 243/243 (99%)	e-146
I51921	pulmonary surfactant-associated protein A1 - human, 248 aa.	1..243 6..248	238/243 (97%) 241/243 (98%)	e-145

P07714	Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A) (PSAP) (Alveolar proteinosis protein) (35 kDa pulmonary surfactant-associated protein) - Homo sapiens (Human), 248 aa.	1..243 6..248	235/243 (96%) 240/243 (98%)	e-143
LNHUPS	pulmonary surfactant protein A precursor (genomic clone) - human, 248 aa.	1..243 6..248	232/243 (95%) 237/243 (97%)	e-141
Q9TT06	Pulmonary surfactant protein A (Pulmonary surfactant-associated protein A) - Ovis aries (Sheep), 248 aa.	1..243 6..248	183/243 (75%) 208/243 (85%)	e-114

PFam analysis indicates that the NOV17a protein contains the domains shown in the Table 17F.

Table 17F. Domain Analysis of NOV17a			
Pfam Domain	NOV17a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Collagen	32..92	34/61 (56%) 49/61 (80%)	0.00019
Xlink	131..158	13/32 (41%) 19/32 (59%)	0.41
lectin_c	139..243	48/125 (38%) 92/125 (74%)	5e-45

5

#### Example 18.

The NOV18 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 18A.

10

Table 18A. NOV18 Sequence Analysis			
	SEQ ID NO: 101	349 bp	
NOV18a, CG97482-01 DNA Sequence	GGTGAGACAAGGAAGAGGATGTCTGAGCTGGAGAAGGCCATGGTGGCCCTCATCGACGTTTCC ACCAATATTCTGGAAGGGAGGGAGACAAGCACAAGCTGAAGAAATCCGAACTCAAGGAGCTCAT CAACAATGAGCTTTCCATTCTTAGAGGAAATCAAAGAGCAGGAGGTTGTGGACAAAGTCATG GAAACACTGGACAATGATGGAGACGGCGAATGTGACTTCCAGGAATTCATGGCCTTTGTGCCA TGGTTACTACTGCCCCGACGAGTCTTTGAACATGAGTGAGATTAGAAAGCAGCCAAACCTTT CCTGTAACAGAGACGGTCATGCAAGAAAG		
	ORF Start: ATG at 19		ORF Stop: TGA at 295
	SEQ ID NO: 102	92 aa	MW at 10766.0kD
NOV18a,	MSELEKAMVALIDVFHQYSGREGDKHKLKKSELKELINNELSHFLEEIKEQEVVDKVMETLDND		

CG97482-01 Protein Sequence	GDGECDFQEFMAFVAMVTTARHEFFFEHE
--------------------------------	-------------------------------

	SEQ ID NO: 103	271 bp	
NOV18b, CG97482-02 DNA Sequence	GGTGAGACAAGGAAGAGGATGTCTGAGCTGGAGAAGGCCATGGTGGCCCTCATCGACGTTTTCC ACCAATATTCTGGAAGGGAGGGAGACAAGCACAAGCTGAAGAAATCCGAAGCTCAAGGAGCTCAT CAACAATGAGCTTTCCCATTTCTTAGAGGAAATCAAAGAGCAGGAGGTTGTGGTTACTACTGCC TGCCACGAGTTCTTTGAACATGAGTGAGATTAGAAAGCAGCCAAACCTTTCTGTAAACAGAGAC GGTCATGCAAGAAAG		
	ORF Start: ATG at 19		ORF Stop: TGA at 217

5

	SEQ ID NO: 104	66 aa	MW at 7772.7kD
NOV18b, CG97482-02 Protein Sequence	MSELEKAMVALIDVFHQYSGREGDKHKLKKSELKELINNELSHFLEEIKEQEVVTTACHEFFE HE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 18B.

10

Table 18B. Comparison of NOV18a against NOV18b.		
Protein Sequence	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV18b	1..92 1..66	65/92 (70%) 65/92 (70%)

Further analysis of the NOV18a protein yielded the following properties shown in Table 18C.

Table 18C. Protein Sequence Properties NOV18a	
SignalP analysis:	No Known Signal Sequence Indicated
PSORT II analysis:	<p>PSG: a new signal peptide prediction method  N-region: length 6; pos.chg 1; neg.chg 2  H-region: length 6; peak value 0.00  PSG score: -4.40</p> <p>GvH: von Heijne's method for signal seq. recognition  GvH score (threshold: -2.1): -8.88  possible cleavage site: between 20 and 21</p> <p>&gt;&gt;&gt; Seems to have no N-terminal signal peptide</p> <p>ALOM: Klein et al's method for TM region allocation</p>

	<p>Init position for calculation: 1 Tentative number of TMS(s) for the threshold 0.5:</p> <p>0</p> <p>number of TMS(s) .. fixed PERIPHERAL Likelihood = 7.32 (at 68) ALOM score: 7.32 (number of TMSs: 0)</p> <p>MITDISC: discrimination of mitochondrial targeting seq R content: 0 Hyd Moment(75): 2.53 Hyd Moment(95): 2.95 G content: 0 D/E content: 2 S/T content: 1 Score: -7.61</p> <p>Gavel: indication of cleavage sites for mitochondrial preseq cleavage site motif not found</p> <p>NUCDISC: discrimination of nuclear localization signals pat4: none pat7: none bipartite: none content of basic residues: 10.9% NLS Score: -0.47</p> <p>KDEL: ER retention motif in the C-terminus: none</p> <p>ER Membrane Retention Signals: none</p> <p>SKL: peroxisomal targeting signal in the C-terminus: none</p> <p>PTS2: 2nd peroxisomal targeting signal: none</p> <p>VAC: possible vacuolar targeting motif: none</p> <p>RNA-binding motif: none</p> <p>Actinin-type actin-binding motif: type 1: none type 2: none</p> <p>NMYR: N-myristoylation pattern : none</p> <p>Prenylation motif: none</p> <p>memYQRL: transport motif from cell surface to Golgi: none</p> <p>Tyrosines in the tail: none</p> <p>Dileucine motif in the tail: none</p> <p>checking 63 PROSITE DNA binding motifs: none</p> <p>checking 71 PROSITE ribosomal protein motifs: none</p> <p>checking 33 PROSITE prokaryotic DNA binding motifs: none</p>
--	--

	<p>discrimination  Indication: cytoplasmic  Reliability: 94.1</p> <p>COIL: Lupas's algorithm to detect coiled-coil regions  total: 0 residues</p> <p>-----</p> <p>Final Results (k = 9/23):</p> <p>56.5 %: cytoplasmic  30.4 %: nuclear  8.7 %: mitochondrial  4.3 %: Golgi</p> <p>&gt;&gt; indication for CG97482-01 is cyt (k=23)</p>
--	---

A search of the NOV18a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 18D.

5

Table 18D. Geneseq Results for NOV18a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB97495	Novel human protein SEQ ID NO: 763 - Homo sapiens, 92 aa. [WO200222660-A2, 21-MAR-2002]	1..92 1..92	91/92 (98%) 91/92 (98%)	3e-47
ABP51390	Human MDDT SEQ ID NO 412 - Homo sapiens, 97 aa. [WO200240715-A2, 23-MAY-2002]	1..92 6..97	89/92 (96%) 90/92 (97%)	3e-46
AAW46607	Human brain protein S100b beta subunit - Homo sapiens, 91 aa. [WO9801471-A1, 15-JAN-1998]	2..92 1..91	84/91 (92%) 87/91 (95%)	4e-43
AAM40258	Human polypeptide SEQ ID NO 3403 - Homo sapiens, 94 aa. [WO200153312-A1, 26-JUL-2001]	2..89 3..90	52/88 (59%) 66/88 (74%)	2e-23
AAB45531	Human S100A1 protein - Homo sapiens, 94 aa. [DE19915485-A1, 19-OCT-2000]	2..89 3..90	52/88 (59%) 66/88 (74%)	2e-23

In a BLAST search of public sequence databases, the NOV18a protein was found to have homology to the proteins shown in the BLASTP data in Table 18E.

Table 18E. Public BLASTP Results for NOV18a				
Protein Accession Number	Protein/Organism/Length	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD35011	Sequence 319 from Patent WO0222660 - Homo sapiens (Human), 92 aa.	1..92 1..92	91/92 (98%) 91/92 (98%)	7e-47
P04271	S-100 protein, beta chain - Homo sapiens (Human), 91 aa.	2..92 1..91	90/91 (98%) 90/91 (98%)	3e-46
A48015	S-100 protein beta chain - mouse, 92 aa.	1..92 1..92	90/92 (97%) 90/92 (97%)	4e-46
A26557	S-100 protein beta chain - rat, 92 aa.	1..92 1..92	89/92 (96%) 90/92 (97%)	8e-46
AAA72205	SYNTHETIC CALCIUM-MODULATED PROTEIN S100-BETA GENE, 5' END - synthetic construct, 92 aa (fragment).	1..92 1..92	88/92 (95%) 91/92 (98%)	1e-45

PFam analysis indicates that the NOV1a protein contains the domains shown in the Table 18F.

5

Table 18F. Domain Analysis of NOV18a			
Pfam Domain	NOV18a Match Region	Identities/ Similarities for the Matched Region	Expect Value
S_100	4..47	28/44 (64%) 41/44 (93%)	3.6e-23
efhand	53..81	9/29 (31%) 25/29 (86%)	0.0012

#### Example B: Sequencing Methodology and Identification of NOVX Clones

1. **GeneCalling™ Technology:** This is a proprietary method of performing differential gene expression profiling between two or more samples developed at CuraGen and described by Shimkets, et al., "Gene expression analysis by transcript profiling coupled to a gene database query". Nature Biotechnology 17:198-803 (1999). cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were

obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then digested with up to as many as 120 pairs of restriction enzymes and pairs of linker-adaptors specific for each pair of restriction enzymes were ligated to the appropriate end. The restriction digestion generates a mixture of unique cDNA gene fragments. Limited PCR amplification is performed with primers homologous to the linker adapter sequence where one primer is biotinylated and the other is fluorescently labeled. The doubly labeled material is isolated and the fluorescently labeled single strand is resolved by capillary gel electrophoresis. A computer algorithm compares the electropherograms from an experimental and control group for each of the restriction digestions. This and additional sequence-derived information is used to predict the identity of each differentially expressed gene fragment using a variety of genetic databases. The identity of the gene fragment is confirmed by additional, gene-specific competitive PCR or by isolation and sequencing of the gene fragment.

2. **SeqCalling™ Technology:** cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

3. **PathCalling™ Technology:** The NOVX nucleic acid sequences are derived by laboratory screening of cDNA library by the two-hybrid approach. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, are sequenced. In silico prediction was based on sequences available in CuraGen

Corporation's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The laboratory screening was performed using the methods summarized below: cDNA libraries were derived from various human samples representing multiple tissue  
5 types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then directionally cloned into the appropriate  
10 two-hybrid vector (Gal4-activation domain (Gal4-AD) fusion). Such cDNA libraries as well as commercially available cDNA libraries from Clontech (Palo Alto, CA) were then transferred from E.coli into a CuraGen Corporation proprietary yeast strain (disclosed in U. S. Patents 6,057,101 and 6,083,693, incorporated herein by reference in their entireties).

Gal4-binding domain (Gal4-BD) fusions of a CuraGen Corporation proprietary  
15 library of human sequences was used to screen multiple Gal4-AD fusion cDNA libraries resulting in the selection of yeast hybrid diploids in each of which the Gal4-AD fusion contains an individual cDNA. Each sample was amplified using the polymerase chain reaction (PCR) using non-specific primers at the cDNA insert boundaries. Such PCR product was sequenced; sequence traces were evaluated manually and edited for  
20 corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly  
25 represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Physical clone: the cDNA fragment derived by the screening procedure, covering the entire open reading frame is, as a recombinant DNA, cloned into pACT2 plasmid  
30 (Clontech) used to make the cDNA library. The recombinant plasmid is inserted into the host and selected by the yeast hybrid diploid generated during the screening procedure by the mating of both CuraGen Corporation proprietary yeast strains N106' and YULH (U. S. Patents 6,057,101 and 6,083,693).

4. **RACE:** Techniques based on the polymerase chain reaction such as rapid amplification of cDNA ends (RACE), were used to isolate or complete the predicted sequence of the cDNA of the invention. Usually multiple clones were sequenced from one or more human samples to derive the sequences for fragments. Various human tissue samples from different donors were used for the RACE reaction. The sequences derived from these procedures were included in the SeqCalling Assembly process described in preceding paragraphs.

5. **Exon Linking:** The NOVX target sequences identified in the present invention were subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The PCR product derived from exon linking was cloned into the pCR2.1 vector from Invitrogen. The resulting bacterial clone has an insert covering the entire open reading frame cloned into the pCR2.1 vector. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

6. **Physical Clone:** Exons were predicted by homology and the intron/exon boundaries were determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed  
5 sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent inconsistencies thereby obtaining the sequences encoding the full-length protein.

The PCR product derived by exon linking, covering the entire open reading frame,  
10 was cloned into the pCR2.1 vector from Invitrogen to provide clones used for expression and screening purposes.

#### **Example C. Quantitative expression analysis of clones in various cells and tissues**

The quantitative expression of various clones was assessed using microtiter plates  
15 containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on an Applied Biosystems ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples  
20 derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an emphasis on metabolic diseases), AI\_comprehensive\_panel (containing normal tissue and samples from autoinflammatory diseases), Panel CNSD.01 (containing samples from  
25 normal and diseased brains) and CNS\_neurodegeneration\_panel (containing samples from normal and Alzheimer's diseased brains).

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would  
30 be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example,  $\beta$ -actin and GAPDH). Normalized RNA (5  $\mu$ l) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix Reagents (Applied Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions.

In other cases, non-normalized RNA samples were converted to single strand cDNA (sscDNA) using Superscript II (Invitrogen Corporation; Catalog No. 18064-147) and random hexamers according to the manufacturer's instructions. Reactions containing up to 10  $\mu$ g of total RNA were performed in a volume of 20  $\mu$ l and incubated for 60 minutes at 42°C. This reaction can be scaled up to 50  $\mu$ g of total RNA in a final volume of 100  $\mu$ l. sscDNA samples are then normalized to reference nucleic acids as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions.

Probes and primers were designed for each assay according to Applied Biosystems Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature ( $T_m$ ) range = 58°-60°C, primer optimal  $T_m$  = 59°C, maximum primer difference = 2°C, probe does not have 5'G, probe  $T_m$  must be 10°C greater than primer  $T_m$ , amplicon size 75bp to 100bp. The probes and primers selected (see below) were synthesized by Synthesgen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900nM each, and probe, 200nM.

PCR conditions: When working with RNA samples, normalized RNA from each tissue and each cell line was spotted in each well of either a 96 well or a 384-well PCR plate (Applied Biosystems). PCR cocktails included either a single gene specific probe and primers set, or two multiplexed probe and primers sets (a set specific for the target clone and another gene-specific set multiplexed with the target probe). PCR reactions were set up using TaqMan® One-Step RT-PCR Master Mix (Applied Biosystems, Catalog No. 4313803) following manufacturer's instructions. Reverse transcription was performed at

48°C for 30 minutes followed by amplification/PCR cycles as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

When working with sscDNA samples, normalized sscDNA was used as described previously for RNA samples. PCR reactions containing one or two sets of probe and primers were set up as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification was performed as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were analyzed and processed as described previously.

#### **Panels 1, 1.1, 1.2, and 1.3D**

The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in these panels are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on these panels are comprised of samples derived from all major organ systems from single adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

In the results for Panels 1, 1.1, 1.2 and 1.3D, the following abbreviations are used:

5       ca. = carcinoma,  
       \* = established from metastasis,  
       met = metastasis,  
       s cell var = small cell variant,  
       non-s = non-sm = non-small,  
       squam = squamous,  
       pl. eff = pl effusion = pleural effusion,  
       glio = glioma,  
 10       astro = astrocytoma, and  
       neuro = neuroblastoma.

#### **General\_screening\_panel\_v1.4, v1.5, v1.6 and 1.7**

The plates for Panels 1.4, 1.5, 1.6 and 1.7 include 2 control wells (genomic DNA control and chemistry control) and 88 to 94 wells containing cDNA from various samples. The samples in Panels 1.4, 1.5, 1.6 and 1.7 are broken into 2 classes: samples derived from
 15   cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in Panels 1.4, 1.5, 1.6 and 1.7 are widely available through the American Type Culture Collection (ATCC), a
 20   repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on Panels 1.4, 1.5, 1.6 and 1.7 are comprised of pools of samples derived from all major organ systems from 2 to 5 different adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver,
 25   adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose. Abbreviations are as described for Panels 1, 1.1, 1.2, and 1.3D.

#### **Panels 2D, 2.2, 2.3 and 2.4**

30       The plates for Panels 2D, 2.2, 2.3 and 2.4 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI) or from Ardaís or Clinomics). The tissues are derived from human malignancies and in cases where
 35   indicated many malignant tissues have "matched margins" obtained from noncancerous

tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologist at NDRI/CHTN/Ardais/Clinomics). Unmatched RNA samples from tissues without malignancy  
5 (normal tissues) were also obtained from Ardais or Clinomics. This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent  
10 tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

#### 15 **HASS Panel v 1.0**

The HASS panel v 1.0 plates are comprised of 93 cDNA samples and two controls. Specifically, 81 of these samples are derived from cultured human cancer cell lines that had been subjected to serum starvation, acidosis and anoxia for different time periods as well as controls for these treatments, 3 samples of human primary cells, 9 samples of malignant  
20 brain cancer (4 medulloblastomas and 5 glioblastomas) and 2 controls. The human cancer cell lines are obtained from ATCC (American Type Culture Collection) and fall into the following tissue groups: breast cancer, prostate cancer, bladder carcinomas, pancreatic cancers and CNS cancer cell lines. These cancer cells are all cultured under standard recommended conditions. The treatments used (serum starvation, acidosis and anoxia) have  
25 been previously published in the scientific literature. The primary human cells were obtained from Clonetics (Walkersville, MD) and were grown in the media and conditions recommended by Clonetics. The malignant brain cancer samples are obtained as part of a collaboration (Henry Ford Cancer Center) and are evaluated by a pathologist prior to CuraGen receiving the samples. RNA was prepared from these samples using the standard  
30 procedures. The genomic and chemistry control wells have been described previously.

#### **ARDAIS Panel v 1.0**

The plates for ARDAIS panel v 1.0 generally include 2 control wells and 22 test samples composed of RNA isolated from human tissue procured by surgeons working in close cooperation with Ardais Corporation. The tissues are derived from human lung malignancies (lung adenocarcinoma or lung squamous cell carcinoma) and in cases where indicated many malignant samples have "matched margins" obtained from noncancerous lung tissue just adjacent to the tumor. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue) in the results below. The tumor tissue and the "matched margins" are evaluated by independent pathologists (the surgical pathologists and again by a pathologist at Ardais). Unmatched malignant and non-malignant RNA samples from lungs were also obtained from Ardais. Additional information from Ardais provides a gross histopathological assessment of tumor differentiation grade and stage. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical state of the patient.

#### **Panel 3D, 3.1 and 3.2**

The plates of Panel 3D, 3.1, and 3.2 are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D, 3.1, 3.2, 1, 1.1., 1.2, 1.3D, 1.4, 1.5, and 1.6 are of the most common cell lines used in the scientific literature.

#### **Panels 4D, 4R, and 4.1D**

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4R) or cDNA (Panels 4D/4.1D) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) was

employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI)  
5 (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville,  
10 MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5ng/ml, TNF alpha at approximately 5-10ng/ml, IFN gamma at approximately 20-50ng/ml, IL-4 at approximately 5-10ng/ml, IL-9 at approximately 5-10ng/ml, IL-13 at  
15 approximately 5-10ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco/Life Technologies,  
20 Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2µg/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at 20-50ng/ml and IL-18 at 5-10ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM  
25 sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5µg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final  
30 concentration of approximately  $2 \times 10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol

( $5.5 \times 10^{-5}$  M) (Gibco), and 10mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions.

- 5 Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10mM Hepes (Gibco), 50ng/ml GMCSF and 5ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of

- 10 (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), 10mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10 $\mu$ g/ml for 6 and 12-14 hours.

- 15 CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive
- 20 selection. CD45RO beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10mM Hepes (Gibco) and plated at  $10^6$  cells/ml onto Falcon 6 well tissue culture plates that
- 25 had been coated overnight with 0.5 $\mu$ g/ml anti-CD28 (Pharmingen) and 3 $\mu$ g/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino
- 30 acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated

6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

5 To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resuspended at  $10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco). To activate the cells, we used PWM at 5µg/ml or anti-CD40  
10 (Pharmingen) at approximately 10µg/ml and IL-4 at 5-10ng/ml. Cells were harvested for RNA preparation at 24, 48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10µg/ml anti-CD28 (Pharmingen) and 2µg/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic  
15 Systems, German Town, MD) were cultured at  $10^5$ - $10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), 10mM Hepes (Gibco) and IL-2 (4ng/ml). IL-12 (5ng/ml) and anti-IL4 (1µg/ml) were used to direct to Th1, while IL-4 (5ng/ml) and anti-IFN gamma (1µg/ml) were used to direct to Th2 and IL-10 at 5ng/ml was used to  
20 direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), 10mM Hepes (Gibco) and IL-2 (1ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as  
25 described above, but with the addition of anti-CD95L (1µg/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound  
30 anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at  $5 \times 10^5$  cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to  $5 \times 10^5$  cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10ng/ml and ionomycin at 1 $\mu$ g/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5ng/ml IL-4, 5ng/ml IL-9, 5ng/ml IL-13 and 25ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately  $10^7$  cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15ml Falcon Tube. An equal volume of isopropanol was added and left at -20°C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300 $\mu$ l of RNase-free water and 35 $\mu$ l buffer (Promega) 5 $\mu$ l DTT, 7 $\mu$ l RNasin and 8 $\mu$ l DNase were added. The tube was incubated at 37°C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNase free water. RNA was stored at -80°C.

#### AI\_comprehensive panel\_v1.0

The plates for AI\_comprehensive panel\_v1.0 include two control wells and 89 test samples comprised of cDNA isolated from surgical and postmortem human tissues obtained from the Backus Hospital and Clinomics (Frederick, MD). Total RNA was

extracted from tissue samples from the Backus Hospital in the Facility at CuraGen. Total RNA from other tissues was obtained from Clinomics.

Joint tissues including synovial fluid, synovium, bone and cartilage were obtained from patients undergoing total knee or hip replacement surgery at the Backus Hospital.

5 Tissue samples were immediately snap frozen in liquid nitrogen to ensure that isolated RNA was of optimal quality and not degraded. Additional samples of osteoarthritis and rheumatoid arthritis joint tissues were obtained from Clinomics. Normal control tissues were supplied by Clinomics and were obtained during autopsy of trauma victims.

Surgical specimens of psoriatic tissues and adjacent matched tissues were provided

10 as total RNA by Clinomics. Two male and two female patients were selected between the ages of 25 and 47. None of the patients were taking prescription drugs at the time samples were isolated.

Surgical specimens of diseased colon from patients with ulcerative colitis and Crohns disease and adjacent matched tissues were obtained from Clinomics. Bowel tissue

15 from three female and three male Crohn's patients between the ages of 41-69 were used. Two patients were not on prescription medication while the others were taking dexamethasone, phenobarbital, or tylenol. Ulcerative colitis tissue was from three male and four female patients. Four of the patients were taking lebvid and two were on phenobarbital.

20 Total RNA from post mortem lung tissue from trauma victims with no disease or with emphysema, asthma or COPD was purchased from Clinomics. Emphysema patients ranged in age from 40-70 and all were smokers, this age range was chosen to focus on patients with cigarette-linked emphysema and to avoid those patients with alpha-1-anti-trypsin deficiencies. Asthma patients ranged in age from 36-75, and excluded

25 smokers to prevent those patients that could also have COPD. COPD patients ranged in age from 35-80 and included both smokers and non-smokers. Most patients were taking corticosteroids, and bronchodilators.

In the labels employed to identify tissues in the AI\_comprehensive panel\_v1.0 panel, the following abbreviations are used:

AI = Autoimmunity  
 Syn = Synovial  
 Normal = No apparent disease  
 Rep22 /Rep20 = individual patients  
 5 RA = Rheumatoid arthritis  
 Backus = From Backus Hospital  
 OA = Osteoarthritis  
 (SS) (BA) (MF) = Individual patients  
 Adj = Adjacent tissue  
 10 Match control = adjacent tissues  
 -M = Male  
 -F = Female  
 COPD = Chronic obstructive pulmonary disease

#### **AL05 chondrosarcoma**

15 The AL05 chondrosarcoma plates are comprised of SW1353 cells that had been subjected to serum starvation and treatment with cytokines that are known to induce MMP (1, 3 and 13) synthesis (eg. IL1beta). These treatments include: IL-1beta (10 ng/ml), IL-1beta + TNF-alpha (50 ng/ml), IL-1beta + Oncostatin (50 ng/ml) and PMA (100 ng/ml). The SW1353 cells were obtained from the ATCC (American Type Culture Collection) and  
 20 were all cultured under standard recommended conditions. The SW1353 cells were plated at  $3 \times 10^5$  cells/ml (in DMEM medium-10 % FBS) in 6-well plates. The treatment was done in triplicate, for 6 and 18 h. The supernatants were collected for analysis of MMP 1, 3 and 13 production and for RNA extraction. RNA was prepared from these samples using the standard procedures.

#### **25 Panels 5D and 5I**

The plates for Panel 5D and 5I include two control wells and a variety of cDNAs isolated from human tissues and cell lines with an emphasis on metabolic diseases. Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study. Cells were obtained during different stages in the differentiation of adipocytes from human  
 30 mesenchymal stem cells. Human pancreatic islets were also obtained.

In the Gestational Diabetes study subjects are young (18 - 40 years), otherwise healthy women with and without gestational diabetes undergoing routine (elective) Caesarean section. After delivery of the infant, when the surgical incisions were being repaired/closed, the obstetrician removed a small sample (<1 cc) of the exposed metabolic  
 35 tissues during the closure of each surgical level. The biopsy material was rinsed in sterile

saline, blotted and fast frozen within 5 minutes from the time of removal. The tissue was then flash frozen in liquid nitrogen and stored, individually, in sterile screw-top tubes and kept on dry ice for shipment to or to be picked up by CuraGen. The metabolic tissues of interest include uterine wall (smooth muscle), visceral adipose, skeletal muscle (rectus) and subcutaneous adipose. Patient descriptions are as follows:

10 Patient 2: Diabetic Hispanic, overweight, not on insulin  
Patient 7-9: Nondiabetic Caucasian and obese (BMI>30)  
Patient 10: Diabetic Hispanic, overweight, on insulin  
Patient 11: Nondiabetic African American and overweight  
Patient 12: Diabetic Hispanic on insulin

Adipocyte differentiation was induced in donor progenitor cells obtained from Osirus (a division of Clonetics/BioWhittaker) in triplicate, except for Donor 3U which had only two replicates. Scientists at Clonetics isolated, grew and differentiated human mesenchymal stem cells (HuMSCs) for CuraGen based on the published protocol found in Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells Science Apr 2 1999: 143-147. Clonetics provided Trizol lysates or frozen pellets suitable for mRNA isolation and ds cDNA production. A general description of each donor is as follows:

20 Donor 2 and 3 U: Mesenchymal Stem cells, Undifferentiated Adipose  
Donor 2 and 3 AM: Adipose, AdiposeMidway Differentiated  
Donor 2 and 3 AD: Adipose, Adipose Differentiated

Human cell lines were generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells, and adrenal cortical adenoma cells. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. All samples were processed at CuraGen to produce single stranded cDNA.

30 Panel 5I contains all samples previously described with the addition of pancreatic islets from a 58 year old female patient obtained from the Diabetes Research Institute at the University of Miami School of Medicine. Islet tissue was processed to total RNA at an outside source and delivered to CuraGen for addition to panel 5I.

In the labels employed to identify tissues in the 5D and 5I panels, the following abbreviations are used:

- 5 GO Adipose = Greater Omentum Adipose
- SK = Skeletal Muscle
- UT = Uterus
- PL = Placenta
- AD = Adipose Differentiated
- AM = Adipose Midway Differentiated
- U = Undifferentiated Stem Cells

#### 10 Panel CNSD.01

- The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

- Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supranuclear Palsy, Depression, and "Normal controls".
- 20 Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus palladus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus palladus, thus this
- 25 region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

- In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:
- 30

PSP = Progressive supranuclear palsy  
Sub Nigra = Substantia nigra  
Glob Palladus= Globus palladus

Temp Pole = Temporal pole  
 Cing Gyr = Cingulate gyrus  
 BA 4 = Brodman Area 4

#### **Panel CNS\_Neurodegeneration\_V1.0**

5           The plates for Panel CNS\_Neurodegeneration\_V1.0 include two control wells and  
 47 test samples comprised of cDNA isolated from postmortem human brain tissue obtained  
 from the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain  
 and Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are  
 removed from calvaria of donors between 4 and 24 hours after death, sectioned by  
 10   neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and  
 examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains six brains  
 from Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who  
 showed no evidence of dementia prior to death. The eight normal control brains are divided  
 15   into two categories: Controls with no dementia and no Alzheimer's like pathology  
 (Controls) and controls with no dementia but evidence of severe Alzheimer's like  
 pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no  
 evidence of plaques, 3 = severe AD senile plaque load). Within each of these brains, the  
 following regions are represented: hippocampus, temporal cortex (Brodman Area 21),  
 20   parietal cortex (Brodman area 7), and occipital cortex (Brodman area 17). These regions  
 were chosen to encompass all levels of neurodegeneration in AD. The hippocampus is a  
 region of early and severe neuronal loss in AD; the temporal cortex is known to show  
 neurodegeneration in AD after the hippocampus; the parietal cortex shows moderate  
 neuronal death in the late stages of the disease; the occipital cortex is spared in AD and  
 25   therefore acts as a "control" region within AD patients. Not all brain regions are  
 represented in all cases.

In the labels employed to identify tissues in the CNS\_Neurodegeneration\_V1.0  
 panel, the following abbreviations are used:

30           AD = Alzheimer's disease brain; patient was demented and showed AD-like  
 pathology upon autopsy  
 Control = Control brains; patient not demented, showing no neuropathology  
 Control (Path) = Control brains; patient not demented but showing severe AD-like  
 pathology

SupTemporal Ctx = Superior Temporal Cortex  
 Inf Temporal Ctx = Inferior Temporal Cortex

**A. CG115907-02 (NOV2d), CG115907-03 (NOV2c), and CG115907-04 (NOV2b): PK- 120.**

- 5 Expression of gene CG115907-02, CG115907-03, and CG115907-04 was assessed using the primer-probe sets Ag6155, Ag6156 and Ag6131, described in Tables AA, AB and AC. Results of the RTQ-PCR runs are shown in Tables AD and AE. Please note that primer-probe set Ag6155 is specific for CG115907-03 and Ag6156 is specific for CG115907-04.

10 **Table AA. Probe Name Ag6155**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -atcttgcttgcttcagcaa-3'	19	2113	105
Probe	TET-5' -caaatcctgatccagctgtgtctcgt-3' -TAMRA	26	2144	106
Reverse	5' -ggatggcagacatattcatgac-3'	22	2170	107

**Table AB. Probe Name Ag6156**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -ggccatcttgcttgctt-3'	17	2109	108
Probe	TET-5' -atcctgatccagctgtgtctcgtgtc-3' -TAMRA	26	2147	109
Reverse	5' -ctccctctcatactgcatattcat-3'	24	2173	110

15

**Table AC. Probe Name Ag6131**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -gtccactcagctggagctg-3'	19	1981	111
Probe	TET-5' -aacttggaactcccaggacctcctgat-3' -TAMRA	26	2045	112
Reverse	5' -cagctggatcaggatttgag-3'	20	2142	113

**Table AD. CNS neurodegeneration v1.0**

20

Tissue Name	Rel. Exp.(%) Ag6131, un 253574594	Tissue Name	Rel. Exp.(%) Ag6131, Run 253574594
-------------	--------------------------------------	-------------	---------------------------------------

AD 1 Hippo	0.0	Control (Path) 3 Temporal Ctx	3.4
AD 2 Hippo	25.0	Control (Path) 4 Temporal Ctx	51.8
AD 3 Hippo	0.0	AD 1 Occipital Ctx	3.8
AD 4 Hippo	11.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	99.3	AD 3 Occipital Ctx	0.0
AD 6 Hippo	55.1	AD 4 Occipital Ctx	10.6
Control 2 Hippo	37.6	AD 5 Occipital Ctx	22.8
Control 4 Hippo	14.3	AD 6 Occipital Ctx	29.7
Control (Path) 3 Hippo	9.0	Control 1 Occipital Ctx	0.0
AD 1 Temporal Ctx	2.6	Control 2 Occipital Ctx	79.6
AD 2 Temporal Ctx	0.0	Control 3 Occipital Ctx	30.8
AD 3 Temporal Ctx	0.0	Control 4 Occipital Ctx	9.0
AD 4 Temporal Ctx	22.1	Control (Path) 1 Occipital Ctx	68.8
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	17.7
AD 5 Sup Temporal Ctx	28.9	Control (Path) 3 Occipital Ctx	5.0
AD 6 Inf Temporal Ctx	56.6	Control (Path) 4 Occipital Ctx	20.6
AD 6 Sup Temporal Ctx	39.2	Control 1 Parietal Ctx	0.0
Control 1 Temporal Ctx	0.0	Control 2 Parietal Ctx	59.0
Control 2 Temporal Ctx	33.0	Control 3 Parietal Ctx	11.1
Control 3 Temporal Ctx	19.9	Control (Path) 1 Parietal Ctx	26.4
Control 3 Temporal Ctx	9.0	Control (Path) 2 Parietal Ctx	13.5
Control (Path) 1 Temporal Ctx	57.0	Control (Path) 3 Parietal Ctx	4.1
Control (Path) 2 Temporal Ctx	55.5	Control (Path) 4 Parietal Ctx	56.3

**Table AE. General screening panel v1.5**

Tissue Name	Rel. Exp.(%) Ag6131, Run 253101092	issue Name	Rel. Exp.(%) Ag6131, Run 253101092
Adipose	0.0	Renal ca. TK-10	0.1
Melanoma* Hs688(A).T	0.0	Bladder	12.3
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.1
Melanoma* M14	0.1	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.1	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.2
Placenta	0.0	Colon cancer tissue	0.1
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0

Ovarian ca. SK-OV-3	0.1	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.1
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.1
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.1
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.1
Ovary	1.3	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.2
Breast ca. MDA-MB-231	0.1	Lymph Node Pool	0.2
Breast ca. BT 549	0.1	Fetal Skeletal Muscle	2.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	1.7
Breast ca. MDA-N	0.0	Spleen Pool	0.1
Breast Pool	0.2	Thymus Pool	0.4
Trachea	0.2	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.2	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.8	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.1	CNS cancer (astro) SNB-75	0.1
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	1.0
Lung ca. A549	0.2	Brain (Amygdala) Pool	0.1
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.4
Lung ca. NCI-H23	0.1	Brain (fetal)	0.2
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.1
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.1	Brain (Substantia nigra) Pool	0.3
Liver	100.0	Brain (Thalamus) Pool	0.2
Fetal Liver	41.2	Brain (whole)	3.6
Liver ca. HepG2	0.0	Spinal Cord Pool	0.1
Kidney Pool	0.7	Adrenal Gland	0.6
Fetal Kidney	0.5	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	2.3

**CNS\_neurodegeneration\_v1.0 Summary:** Ag6131 Low levels of expression of this gene is detected in the brains from control and Alzheimer's patients. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Ag6155/Ag6156 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

**General\_screening\_panel\_v1.5 Summary:** Ag6131 Highest expression of this gene is detected in liver (CT=25.2). High expression of this gene is mainly seen in adult and fetal liver, with moderate to low levels of expression in adult and fetal skeletal muscle, adrenal gland and pancreas. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, moderate to low expression of this gene is also seen in whole brain, fetal brain, substantia nigra, thalamus, cerebellum, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Low expression of this gene is also seen in a brain cancer SF-295 cell line. Therefore, therapeutic modulation of this gene may be useful in the treatment of brain cancer.

Ag6155/Ag6156 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

Ag6155/Ag6156 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

#### **B. CG139008-01 (NOV3a): novel secreted.**

Expression of gene CG139008-01 was assessed using the primer-probe sets Ag243 and Ag7477, described in Tables BA and BB. Results of the RTQ-PCR runs are shown in Tables BC and BD.

#### **25 Table BA. Probe Name Ag243**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - caagggcatcaccaatttga - 3'	20	186	114
Probe	TET - 5' - aggatgtccagctgcccgatca - 3' - TAMRA	24	212	115

Reverse	5'-gcccaactccaggtacaaagttc-3'	22	240	116
---------	-------------------------------	----	-----	-----

**Table BB. Probe Name Ag7477**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ttcttttgacagattactgagctt-3'	24	1134	117
Probe	TET-5'-tcctcatcgattggcaacttcaatga-3'-TAMRA	26	1168	118
Reverse	5'-tcttcgagatagctggtgatg-3'	21	1212	119

**5 Table BC. Panel 1.3D**

Tissue Name	el. Exp.(%) Ag243, Run 156536275	Tissue Name	Rel. Exp.(%) Ag243, Run 156536275
Liver adenocarcinoma	0.0	Kidney (fetal)	0.0
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	17.6	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	0.0
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	46.3
Brain (cerebellum)	0.0	Lung	0.0
Brain (hippocampus)	0.0	Lung (fetal)	0.0
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.0	Lung ca. (large cell) NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.0	Mammary gland	0.0
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	0.0
Heart	0.0	Breast ca. BT-549	0.0

Skeletal muscle (fetal)	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.0	Ovarian ca. OVCAR-4	0.0
Spleen	0.0	Ovarian ca. OVCAR-5	0.0
Lymph node	0.0	Ovarian ca. OVCAR-8	0.0
Colorectal	0.0	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	0.0
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	100.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	0.0

Table BD. Panel 2D

	Rel. Exp.(%) Ag243, Run 156536477	Tissue Name	Rel. Exp.(%) Ag243, Run 156536477
Normal Colon	0.0	Kidney Margin 8120608	0.0
CC Well to Mod Diff (ODO3866)	0.0	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	0.0	Kidney Margin 8120614	0.0
CC Gr.2 rectosigmoid (ODO3868)	0.0	Kidney Cancer 9010320	0.0
CC Margin (ODO3868)	0.0	Kidney Margin 9010321	0.0
CC Mod Diff (ODO3920)	0.0	Normal Uterus	0.0
CC Margin (ODO3920)	0.0	Uterus Cancer 064011	0.0
CC Gr.2 ascend colon (ODO3921)	0.0	Normal Thyroid	0.0
CC Margin (ODO3921)	0.0	Thyroid Cancer 064010	0.0
CC from Partial Hepatectomy (ODO4309) Mets	0.0	Thyroid Cancer A302152	0.0
Liver Margin (ODO4309)	0.0	Thyroid Margin A302153	0.0
Colon mets to lung (OD04451-01)	0.0	Normal Breast	0.0
Lung Margin (OD04451-02)	0.0	Breast Cancer (OD04566)	0.0
Normal Prostate 6546-1	0.0	Breast Cancer (OD04590-01)	0.0
Prostate Cancer (OD04410)	21.9	Breast Cancer Mets (OD04590-03)	0.0

Prostate Margin (OD04410)	0.0	Breast Cancer Metastasis (OD04655-05)	0.0
Prostate Cancer (OD04720-01)	0.0	Breast Cancer 064006	0.0
Prostate Margin (OD04720-02)	0.0	Breast Cancer 1024	0.0
Normal Lung 061010	0.0	Breast Cancer 9100266	100.0
Lung Met to Muscle (ODO4286)	0.0	Breast Margin 9100265	0.0
Muscle Margin (ODO4286)	0.0	Breast Cancer A209073	0.0
Lung Malignant Cancer (OD03126)	15.8	Breast Margin A209073	0.0
Lung Margin (OD03126)	0.0	Normal Liver	0.0
Lung Cancer (OD04404)	0.0	Liver Cancer 064003	0.0
Lung Margin (OD04404)	0.0	Liver Cancer 1025	0.0
Lung Cancer (OD04565)	0.0	Liver Cancer 1026	21.9
Lung Margin (OD04565)	0.0	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237-01)	0.0	Liver Tissue 6004-N	0.0
Lung Margin (OD04237-02)	0.0	Liver Cancer 6005-T	0.0
Ocular Mel Met to Liver (ODO4310)	0.0	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.0	Normal Bladder	0.0
Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer 1023	0.0
Lung Margin (OD04321)	0.0	Bladder Cancer A302173	0.0
Normal Kidney	0.0	Bladder Cancer (OD04718-01)	0.0
Kidney Ca, Nuclear grade 2 (OD04338)	0.0	Bladder Normal Adjacent (OD04718-03)	0.0
Kidney Margin (OD04338)	0.0	Normal Ovary	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Ovarian Cancer 064008	0.0
Kidney Margin (OD04339)	0.0	Ovarian Cancer (OD04768-07)	0.0
Kidney Ca, Clear cell type (OD04340)	0.0	Ovary Margin (OD04768-08)	0.0
Kidney Margin (OD04340)	0.0	Normal Stomach	0.0
Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 9060358	0.0
Kidney Margin (OD04348)	0.0	Stomach Margin 9060359	0.0
Kidney Cancer (OD04622-01)	0.0	Gastric Cancer 9060395	0.0
Kidney Margin (OD04622-03)	0.0	Stomach Margin 9060394	0.0
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	0.0
Kidney Margin (OD04450-03)	0.0	Stomach Margin 9060396	0.0
Kidney Cancer 8120607	0.0	Gastric Cancer 064005	0.0

**Panel 1 Summary:** Ag243 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

**Panel 1.3D Summary:** Ag243 Expression of this gene is restricted to the trachea and a liver cancer cell line (CTs=33.5 - 34.5). Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to

detect the presence of liver cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of liver cancer.

**Panel 2D Summary:** Ag243 Expression of this gene is restricted to the a breast cancer cell line (CT=34.5). Thus, expression of this gene could be used to differentiate  
5 between this sample and other samples on this panel and as a marker to detect the presence of breast cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of breast cancer.

**Panel 4.1D Summary:** Ag7477 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

10 **Panel 4D Summary:** Ag243 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

### C. CG165528-01 (NOV9a): Neurexin I alpha precursor.

15 Expression of gene CG165528-01 was assessed using the primer-probe set Ag5964, described in Table CA. Results of the RTQ-PCR runs are shown in Tables CB and CC.

**Table CA. Probe Name Ag5964**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -gatgtgaaagtcaccaggaatct-3'	23	1204	120
Probe	TET-5' -ttaccatagcgtgtccaatgcctgag-3' -TAMRA	26	1236	121
Reverse	5' -gatattgtcaccgaacaatgtagttt-3'	26	1264	122

20 **Table CB. CNS neurodegeneration v1.0**

Tissue Name	Rel. Exp.(%) Ag5964, Run 248162714	Rel. Exp.(%) Ag5964, Ru 268784143	Tissue Name	Rel. Exp.(%) Ag5964, Run 248162714	Rel. Exp.(%) Ag5964, Run 268784143
AD 1 Hippo	8.6	11.5	Control (Path) 3 Temporal Ctx	17.6	16.6

AD 2 Hippo	33.7	35.8	Control (Path) 4 Temporal Ctx	62.4	47.3
AD 3 Hippo	3.5	7.7	AD 1 Occipital Ctx	11.0	13.1
AD 4 Hippo	15.0	12.6	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	100.0	100.0	AD 3 Occipital Ctx	3.9	1.1
AD 6 Hippo	49.7	36.6	AD 4 Occipital Ctx	34.6	18.8
Control 2 Hippo	30.6	24.0	AD 5 Occipital Ctx	41.5	29.3
Control 4 Hippo	19.9	17.8	AD 6 Occipital Ctx	41.5	41.8
Control (Path) 3 Hippo	11.0	6.6	Control 1 Occipital Ctx	8.5	2.2
AD 1 Temporal Ctx	11.4	9.5	Control 2 Occipital Ctx	66.0	48.6
AD 2 Temporal Ctx	40.6	29.3	Control 3 Occipital Ctx	28.5	15.4
AD 3 Temporal Ctx	4.6	2.9	Control 4 Occipital Ctx	7.7	13.4
AD 4 Temporal Ctx	37.9	27.9	Control (Path) 1 Occipital Ctx	92.0	86.5
AD 5 Inf Temporal Ctx	97.3	59.9	Control (Path) 2 Occipital Ctx	21.6	12.9
AD 5 Sup Temporal Ctx	45.1	29.3	Control (Path) 3 Occipital Ctx	5.8	4.9
AD 6 Inf Temporal Ctx	49.3	55.9	Control (Path) 4 Occipital Ctx	19.8	10.9
AD 6 Sup Temporal Ctx	57.4	75.3	Control 1 Parietal Ctx	22.5	12.9
Control 1 Temporal Ctx	28.7	18.2	Control 2 Parietal Ctx	40.9	24.1
Control 2 Temporal Ctx	47.0	21.5	Control 3 Parietal Ctx	20.4	19.1
Control 3 Temporal Ctx	32.3	24.1	Control (Path) 1 Parietal Ctx	95.9	72.2
Control 4 Temporal Ctx	14.8	17.2	Control (Path) 2 Parietal Ctx	32.3	31.4
Control (Path) 1 Temporal Ctx	80.7	97.9	Control (Path) 3 Parietal Ctx	4.9	12.2
Control (Path) 2 Temporal Ctx	57.8	49.3	Control (Path) 4 Parietal Ctx	66.4	33.0

Table CC. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag5964, Run 248163367	issue Name	Rel. Exp.(%) Ag5964, Run 248163367
-------------	--	------------	--

Adipose	3.6	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	8.1
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	3.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	2.5	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	5.7	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.3
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	13.8
Ovarian ca. IGROV-1	0.0	Stomach Pool	4.5
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	4.3
Ovary	0.0	Fetal Heart	0.5
Breast ca. MCF-7	0.0	Heart Pool	4.2
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	2.8
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	5.4
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.7
Breast ca. MDA-N	0.0	Spleen Pool	1.0
Breast Pool	0.0	Thymus Pool	1.3
Trachea	5.1	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	7.0	CNS cancer (neuro;met) SK-N-AS	0.4
Lung ca. NCI-N417	4.2	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	4.8	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	3.2	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	37.4
Lung ca. NCI-H526	0.0	Brain (cerebellum)	43.8
Lung ca. NCI-H23	0.0	Brain (fetal)	100.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	46.3
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	47.6
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	35.4
Liver	0.0	Brain (Thalamus) Pool	61.1
Fetal Liver	0.7	Brain (whole)	46.3
Liver ca. HepG2	0.0	Spinal Cord Pool	20.9
Kidney Pool	2.0	Adrenal Gland	2.4
Fetal Kidney	5.6	Pituitary gland Pool	7.6
Renal ca. 786-0	0.0	Salivary Gland	2.3

Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	1.3

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5964 Two experiments with same probe-primer sets are in good agreement. This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. This gene is found to be slightly down-regulated in the temporal cortex of Alzheimer's disease patients. Therefore, up-regulation of this gene or its protein product, or treatment with specific agonists for this receptor may be of use in reversing the dementia/memory loss associated with this disease and neuronal death.

**General\_screening\_panel\_v1.5 Summary:** Ag5964 Expression of this gene is seen exclusively in all the regions of brain region, with highest expression in fetal brain (CT=30.9). Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

#### D. CG165528-02 (NOV9b): Neurexin I beta.

Expression of gene CG165528-02 was assessed using the primer-probe set Ag7944, described in Table DA. Results of the RTQ-PCR runs are shown in Table DB.

**Table DA. Probe Name Ag7944**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gcaccacatccaccatttc-3'	19	213	123
Probe	TET-5'-cagcagcaagcatcattcagtgcc-3'-TAMRA	24	237	124
Reverse	5'-gatgccggtgacctgtaga-3'	19	269	125

**Table DB. CNS neurodegeneration v1.0**

Tissue Name	Rel. Exp.(%) Ag7944, Run 319510463	Tissue Name	Rel. Exp.(%) Ag7944, Run 319510463
AD 1 Hippo	5.9	Control (Path) 3 Temporal Ctx	4.7
AD 2 Hippo	13.2	Control (Path) 4 Temporal Ctx	48.3

AD 3 Hippo	3.5	AD 1 Occipital Ctx	17.4
AD 4 Hippo	4.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	3.3
AD 6 Hippo	31.6	AD 4 Occipital Ctx	23.8
Control 2 Hippo	15.9	AD 5 Occipital Ctx	56.6
Control 4 Hippo	3.8	AD 6 Occipital Ctx	16.2
Control (Path) 3 Hippo	2.4	Control 1 Occipital Ctx	1.7
AD 1 Temporal Ctx	9.2	Control 2 Occipital Ctx	64.6
AD 2 Temporal Ctx	31.4	Control 3 Occipital Ctx	22.7
AD 3 Temporal Ctx	6.4	Control 4 Occipital Ctx	2.4
AD 4 Temporal Ctx	23.8	Control (Path) 1 Occipital Ctx	72.2
AD 5 Inf Temporal Ctx	79.0	Control (Path) 2 Occipital Ctx	12.9
AD 5 Sup Temporal Ctx	20.0	Control (Path) 3 Occipital Ctx	0.9
AD 6 Inf Temporal Ctx	30.8	Control (Path) 4 Occipital Ctx	23.3
AD 6 Sup Temporal Ctx	40.3	Control 1 Parietal Ctx	5.8
Control 1 Temporal Ctx	2.7	Control 2 Parietal Ctx	37.4
Control 2 Temporal Ctx	33.4	Control 3 Parietal Ctx	16.6
Control 3 Temporal Ctx	19.8	Control (Path) 1 Parietal Ctx	81.8
Control 3 Temporal Ctx	7.7	Control (Path) 2 Parietal Ctx	28.1
Control (Path) 1 Temporal Ctx	44.8	Control (Path) 3 Parietal Ctx	1.8
Control (Path) 2 Temporal Ctx	47.6	Control (Path) 4 Parietal Ctx	62.4

**CNS\_neurodegeneration\_v1.0 Summary:** Ag7944 No differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. However, this panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

**Panel 4.1D Summary:** Ag7944 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

#### **E. CG165666-01 (NOV10a): CGI-87 PROTEIN.**

Expression of gene CG165666-01 was assessed using the primer-probe set Ag5963, described in Table EA. Results of the RTQ-PCR runs are shown in Tables EB, EC and ED.

**Table EA. Probe Name Ag5963**

Primer s	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gagacaagtcctaaatgccgac-3'	22	159	126
Probe	TET-5'-aacaatcttttgttgattgaaacagctaatacct-3'- TAMRA	34	188	127
Reverse	5'-ctagtagtgccagcctgacaaa-3'	22	226	128

**Table EB. CNS neurodegeneration v1.0**

5

Tissue Name	Rel. Exp.(%) Ag5963, Run 248162713	Tissue Name	Rel. Exp.(%) Ag5963, Run 248162713
AD 1 Hippo	10.7	Control (Path) 3 Temporal Ctx	3.8
AD 2 Hippo	28.7	Control (Path) 4 Temporal Ctx	27.5
AD 3 Hippo	7.5	AD 1 Occipital Ctx	12.6
AD 4 Hippo	7.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	6.4
AD 6 Hippo	50.0	AD 4 Occipital Ctx	16.8
Control 2 Hippo	29.7	AD 5 Occipital Ctx	54.3
Control 4 Hippo	13.6	AD 6 Occipital Ctx	16.4
Control (Path) 3 Hippo	4.8	Control 1 Occipital Ctx	3.3
AD 1 Temporal Ctx	27.4	Control 2 Occipital Ctx	57.4
AD 2 Temporal Ctx	39.0	Control 3 Occipital Ctx	13.7
AD 3 Temporal Ctx	4.2	Control 4 Occipital Ctx	6.3
AD 4 Temporal Ctx	23.2	Control (Path) 1 Occipital Ctx	84.1
AD 5 Inf Temporal Ctx	72.7	Control (Path) 2 Occipital Ctx	11.0
AD 5 Sup Temporal Ctx	40.6	Control (Path) 3 Occipital Ctx	3.0
AD 6 Inf Temporal Ctx	46.3	Control (Path) 4 Occipital Ctx	21.8
AD 6 Sup Temporal Ctx	42.6	Control 1 Parietal Ctx	8.0
Control 1 Temporal Ctx	5.4	Control 2 Parietal Ctx	25.3
Control 2 Temporal Ctx	29.9	Control 3 Parietal Ctx	14.5
Control 3 Temporal Ctx	9.3	Control (Path) 1 Parietal Ctx	63.3
Control 3 Temporal Ctx	8.7	Control (Path) 2 Parietal Ctx	24.8
Control (Path) 1 Temporal Ctx	51.4	Control (Path) 3 Parietal Ctx	1.9
Control (Path) 2 Temporal Ctx	36.6	Control (Path) 4 Parietal Ctx	34.9

Table EC. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag5963, Run 247945158	issue Name	Rel. Exp.(%) Ag5963, Run 247945158
Adipose	9.9	Renal ca. TK-10	21.2
Melanoma* Hs688(A).T	56.6	Bladder	8.8
Melanoma* Hs688(B).T	30.4	Gastric ca. (liver met.) NCI-N87	100.0
Melanoma* M14	15.7	Gastric ca. KATO III	84.1
Melanoma* LOXIMVI	60.7	Colon ca. SW-948	9.1
Melanoma* SK-MEL-5	20.9	Colon ca. SW480	24.8
Squamous cell carcinoma SCC-4	12.6	Colon ca.* (SW480 met) SW620	19.5
Testis Pool	15.3	Colon ca. HT29	13.6
Prostate ca.* (bone met) PC-3	42.9	Colon ca. HCT-116	65.5
Prostate Pool	24.5	Colon ca. CaCo-2	23.0
Placenta	5.3	Colon cancer tissue	23.0
Uterus Pool	14.2	Colon ca. SW1116	5.2
Ovarian ca. OVCAR-3	49.3	Colon ca. Colo-205	16.7
Ovarian ca. SK-OV-3	7.8	Colon ca. SW-48	6.1
Ovarian ca. OVCAR-4	6.9	Colon Pool	8.5
Ovarian ca. OVCAR-5	26.1	Small Intestine Pool	9.9
Ovarian ca. IGROV-1	28.9	Stomach Pool	8.0
Ovarian ca. OVCAR-8	10.1	Bone Marrow Pool	7.8
Ovary	5.2	Fetal Heart	13.9
Breast ca. MCF-7	26.6	Heart Pool	17.7
Breast ca. MDA-MB-231	31.9	Lymph Node Pool	0.0
Breast ca. BT 549	16.8	Fetal Skeletal Muscle	5.0
Breast ca. T47D	4.3	Skeletal Muscle Pool	50.0
Breast ca. MDA-N	13.5	Spleen Pool	17.1
Breast Pool	26.4	Thymus Pool	13.0
Trachea	16.2	CNS cancer (glio/astro) U87-MG	27.9
Lung	0.0	CNS cancer (glio/astro) U-118-MG	88.9
Fetal Lung	47.0	CNS cancer (neuro;met) SK-N-AS	72.7
Lung ca. NCI-N417	2.9	CNS cancer (astro) SF-539	38.7
Lung ca. LX-1	35.6	CNS cancer (astro) SNB-75	77.4
Lung ca. NCI-H146	8.8	CNS cancer (glio) SNB-19	9.7
Lung ca. SHP-77	61.1	CNS cancer (glio) SF-295	70.2
Lung ca. A549	95.3	Brain (Amygdala) Pool	29.5
Lung ca. NCI-H526	3.0	Brain (cerebellum)	47.3
Lung ca. NCI-H23	34.9	Brain (fetal)	41.8
Lung ca. NCI-H460	21.0	Brain (Hippocampus) Pool	6.2
Lung ca. HOP-62	9.0	Cerebral Cortex Pool	33.2
Lung ca. NCI-H522	15.1	Brain (Substantia nigra) Pool	10.2

Liver	2.2	Brain (Thalamus) Pool	13.3
Fetal Liver	10.2	Brain (whole)	12.1
Liver ca. HepG2	15.6	Spinal Cord Pool	6.4
Kidney Pool	14.1	Adrenal Gland	6.2
Fetal Kidney	23.0	Pituitary gland Pool	2.1
Renal ca. 786-0	12.1	Salivary Gland	25.3
Renal ca. A498	6.6	Thyroid (female)	5.7
Renal ca. ACHN	42.3	Pancreatic ca. CAPAN2	14.3
Renal ca. UO-31	3.8	Pancreas Pool	30.6

**Table ED. Panel 4.1D**

Tissue Name	Rel. Exp.(%) g5963, Run 247851482	Tissue Name	Rel. Exp.(%) Ag5963, Run 247851482
Secondary Th1 act	47.0	HUVEC IL-1beta	27.9
Secondary Th2 act	55.5	HUVEC IFN gamma	31.0
Secondary Tr1 act	15.3	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	9.5
Secondary Tr1 rest	0.0	Lung Microvascular EC none	84.7
Primary Th1 act	4.8	Lung Microvascular EC TNFalpha + IL-1beta	17.4
Primary Th2 act	34.6	Microvascular Dermal EC none	0.0
Primary Tr1 act	29.5	Microvascular Dermal EC TNFalpha + IL-1beta	5.5
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	22.1
Primary Th2 rest	11.4	Small airway epithelium none	53.2
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	54.0
CD45RA CD4 lymphocyte act	45.1	Coronary artery SMC rest	30.1
CD45RO CD4 lymphocyte act	55.5	Coronary artery SMC TNFalpha + IL-1beta	36.1
CD8 lymphocyte act	2.8	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	27.2	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	24.1
CD4 lymphocyte none	2.7	KU-812 (Basophil) PMA/ionomycin	47.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	32.5
LAK cells rest	13.5	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	35.6
LAK cells IL-2	9.0	Liver cirrhosis	10.0

LAK cells IL-2+IL-12	0.0	NCI-H292 none	20.9
LAK cells IL-2+IFN gamma	12.6	NCI-H292 IL-4	39.0
LAK cells IL-2+ IL-18	12.3	NCI-H292 IL-9	47.0
LAK cells PMA/ionomycin	13.6	NCI-H292 IL-13	69.3
NK Cells IL-2 rest	48.3	NCI-H292 IFN gamma	17.4
Two Way MLR 3 day	10.9	HPAEC none	3.8
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	34.4
Two Way MLR 7 day	0.0	Lung fibroblast none	31.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	33.0
PBMC PWM	5.5	Lung fibroblast IL-4	12.4
PBMC PHA-L	5.8	Lung fibroblast IL-9	21.6
Ramos (B cell) none	3.9	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	21.6	Lung fibroblast IFN gamma	42.9
B lymphocytes PWM	40.3	Dermal fibroblast CCD1070 rest	69.7
B lymphocytes CD40L and IL-4	57.0	Dermal fibroblast CCD1070 TNF alpha	100.0
EOL-1 dbcAMP	43.2	Dermal fibroblast CCD1070 IL-1 beta	37.1
EOL-1 dbcAMP PMA/ionomycin	6.5	Dermal fibroblast IFN gamma	6.7
Dendritic cells none	11.8	Dermal fibroblast IL-4	32.8
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	39.0
Dendritic cells anti-CD40	8.7	Neutrophils TNFa+LPS	0.0
Monocytes rest	6.4	Neutrophils rest	24.8
Monocytes LPS	21.3	Colon	0.0
Macrophages rest	12.5	Lung	0.0
Macrophages LPS	0.0	Thymus	5.1
HUVEC none	21.3	Kidney	23.3
HUVEC starved	41.2		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5963 This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. This gene is found to be slightly upregulated in the temporal cortex of Alzheimer's disease patients. Blockade of this receptor may be of use in the treatment of this disease and decrease neuronal death.

**General\_screening\_panel\_v1.5 Summary:** Ag5963 Highest expression of this gene is detected in a gastric cancer NCI-N87 cell line (CT=31). Moderate to low levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression

or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate to low levels in pancreas, adipose, skeletal muscle, heart, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at moderate to low levels in all regions of the central nervous system examined, including amygdala, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

**Panel 4.1D Summary:** Ag5963 Low expression of this gene is detected in TNF alpha activated dermal fibroblast (CT=34.6). Therefore, therapeutic modulation of this gene may be useful in the treatment of skin disorders, including psoriasis.

#### 20 **F. CG165676-01(NOV11a): INTEGRIN ALPHA-2 PRECURSOR.**

Expression of gene CG165676-01 was assessed using the primer-probe set Ag4510, described in Table FA. Results of the RTQ-PCR runs are shown in Tables FB, FC and FD.

**Table FA. Probe Name Ag4510**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aaaatttcaggcacaccaaag-3'	21	3018	129
Probe	TET-5'-aattgaactgcagaactgcttcctgt-3'-TAMRA	26	3039	130
Reverse	5'-tctcctttcatgtgaacgtctt-3'	22	3087	131

25

**Table FB. AI comprehensive panel v1.0**

Tissue Name	Rel. Exp.(%) Ag4510, Run 46953623	Tissue Name	Rel. Exp.(%) Ag4510, Run 246953623
110967 COPD-F	5.8	112427 Match Control Psoriasis-F	38.2
110980 COPD-F	8.8	112418 Psoriasis-M	6.4
110968 COPD-M	3.8	112723 Match Control Psoriasis-M	5.3
110977 COPD-M	29.7	112419 Psoriasis-M	6.4
110989 Emphysema-F	37.9	112424 Match Control Psoriasis-M	7.3
110992 Emphysema-F	8.4	112420 Psoriasis-M	25.9
110993 Emphysema-F	5.4	112425 Match Control Psoriasis-M	35.6
110994 Emphysema-F	2.1	104689 (MF) OA Bone-Backus	82.4
110995 Emphysema-F	22.4	104690 (MF) Adj "Normal" Bone-Backus	16.6
110996 Emphysema-F	6.7	104691 (MF) OA Synovium-Backus	19.5
110997 Asthma-M	8.2	104692 (BA) OA Cartilage-Backus	1.7
111001 Asthma-F	16.6	104694 (BA) OA Bone-Backus	100.0
111002 Asthma-F	26.6	104695 (BA) Adj "Normal" Bone-Backus	28.9
111003 Atopic Asthma-F	32.8	104696 (BA) OA Synovium-Backus	12.8
111004 Atopic Asthma-F	32.5	104700 (SS) OA Bone-Backus	15.4
111005 Atopic Asthma-F	24.0	104701 (SS) Adj "Normal" Bone-Backus	19.6
111006 Atopic Asthma-F	5.1	104702 (SS) OA Synovium-Backus	32.8
111417 Allergy-M	14.8	117093 OA Cartilage Rep7	17.2
112347 Allergy-M	0.0	112672 OA Bone5	16.8
112349 Normal Lung-F	0.0	112673 OA Synovium5	11.8
112357 Normal Lung-F	16.3	112674 OA Synovial Fluid cells5	8.0
112354 Normal Lung-M	5.7	117100 OA Cartilage Rep14	0.8
112374 Crohns-F	5.3	112756 OA Bone9	30.8
112389 Match Control Crohns-F	31.6	112757 OA Synovium9	3.2
112375 Crohns-F	5.4	112758 OA Synovial Fluid Cells9	8.1
112732 Match Control Crohns-F	22.7	117125 RA Cartilage Rep2	9.9
112725 Crohns-M	3.3	113492 Bone2 RA	50.0
112387 Match Control Crohns-M	5.4	113493 Synovium2 RA	16.6
112378 Crohns-M	0.0	113494 Syn Fluid Cells RA	25.5
112390 Match Control Crohns-M	60.3	113499 Cartilage4 RA	28.5
112726 Crohns-M	27.9	113500 Bone4 RA	42.6
112731 Match Control Crohns-M	12.0	113501 Synovium4 RA	30.1
112380 Ulcer Col-F	23.3	113502 Syn Fluid Cells4 RA	15.9
112734 Match Control Ulcer Col-F	65.5	113495 Cartilage3 RA	25.2
112384 Ulcer Col-F	63.3	113496 Bone3 RA	34.4

112737 Match Control Ulcer Col-F	13.2	113497 Synovium3 RA	17.4
112386 Ulcer Col-F	2.0	113498 Syn Fluid Cells3 RA	45.1
112738 Match Control Ulcer Col-F	25.3	117106 Normal Cartilage Rep20	1.5
112381 Ulcer Col-M	0.4	113663 Bone3 Normal	0.1
112735 Match Control Ulcer Col-M	2.1	113664 Synovium3 Normal	0.0
112382 Ulcer Col-M	35.6	113665 Syn Fluid Cells3 Normal	0.0
112394 Match Control Ulcer Col-M	1.2	117107 Normal Cartilage Rep22	5.6
112383 Ulcer Col-M	39.0	113667 Bone4 Normal	7.5
112736 Match Control Ulcer Col-M	16.3	113668 Synovium4 Normal	6.0
112423 Psoriasis-F	12.3	113669 Syn Fluid Cells4 Normal	9.1

**Table FC. General screening panel v1.4**

Tissue Name	Rel. Exp.(%) Ag4510, Run 222695870	issue Name	Rel. Exp.(%) Ag4510, Run 222695870
Adipose	1.7	Renal ca. TK-10	9.2
Melanoma* Hs688(A).T	0.9	Bladder	8.0
Melanoma* Hs688(B).T	4.3	Gastric ca. (liver met.) NCI-N87	35.1
Melanoma* M14	5.6	Gastric ca. KATO III	21.0
Melanoma* LOXIMVI	40.3	Colon ca. SW-948	3.5
Melanoma* SK-MEL-5	4.3	Colon ca. SW480	12.8
Squamous cell carcinoma SCC-4	22.1	Colon ca.* (SW480 met) SW620	5.6
Testis Pool	0.9	Colon ca. HT29	2.9
Prostate ca.* (bone met) PC-3	36.3	Colon ca. HCT-116	10.4
Prostate Pool	2.3	Colon ca. CaCo-2	5.7
Placenta	0.2	Colon cancer tissue	17.0
Uterus Pool	0.4	Colon ca. SW1116	1.5
Ovarian ca. OVCAR-3	1.4	Colon ca. Colo-205	2.7
Ovarian ca. SK-OV-3	2.2	Colon ca. SW-48	3.5
Ovarian ca. OVCAR-4	0.7	Colon Pool	1.6
Ovarian ca. OVCAR-5	10.4	Small Intestine Pool	1.9
Ovarian ca. IGROV-1	7.4	Stomach Pool	2.3
Ovarian ca. OVCAR-8	1.2	Bone Marrow Pool	0.9
Ovary	0.8	Fetal Heart	1.1
Breast ca. MCF-7	19.3	Heart Pool	0.4
Breast ca. MDA-MB-231	64.6	Lymph Node Pool	1.5
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	1.4

Breast ca. T47D	17.0	Skeletal Muscle Pool	0.1
Breast ca. MDA-N	3.2	Spleen Pool	4.2
Breast Pool	1.8	Thymus Pool	2.0
Trachea	6.6	CNS cancer (glio/astro) U87-MG	23.3
Lung	0.2	CNS cancer (glio/astro) U-118-MG	18.7
Fetal Lung	7.2	CNS cancer (neuro;met) SK-N-AS	24.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.2
Lung ca. LX-1	6.0	CNS cancer (astro) SNB-75	1.9
Lung ca. NCI-H146	0.8	CNS cancer (glio) SNB-19	6.9
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	100.0
Lung ca. A549	10.8	Brain (Amygdala) Pool	1.9
Lung ca. NCI-H526	0.2	Brain (cerebellum)	0.1
Lung ca. NCI-H23	9.9	Brain (fetal)	0.8
Lung ca. NCI-H460	1.0	Brain (Hippocampus) Pool	1.4
Lung ca. HOP-62	17.9	Cerebral Cortex Pool	1.6
Lung ca. NCI-H522	0.1	Brain (Substantia nigra) Pool	2.0
Liver	0.0	Brain (Thalamus) Pool	2.3
Fetal Liver	0.6	Brain (whole)	0.4
Liver ca. HepG2	16.7	Spinal Cord Pool	2.4
Kidney Pool	1.7	Adrenal Gland	3.7
Fetal Kidney	7.1	Pituitary gland Pool	0.2
Renal ca. 786-0	1.1	Salivary Gland	0.7
Renal ca. A498	2.0	Thyroid (female)	1.0
Renal ca. ACHN	1.0	Pancreatic ca. CAPAN2	34.2
Renal ca. UO-31	8.5	Pancreas Pool	1.9

**Table FD. Panel 4.1D**

Tissue Name	Rel. Exp.(%) g4510, Run 246789401	Tissue Name	Rel. Exp.(%) Ag4510, Run 246789401
Secondary Th1 act	9.8	HUVEC IL-1beta	36.1
Secondary Th2 act	10.7	HUVEC IFN gamma	22.8
Secondary Tr1 act	2.4	HUVEC TNF alpha + IFN gamma	3.1
Secondary Th1 rest	0.1	HUVEC TNF alpha + IL4	2.5
Secondary Th2 rest	0.2	HUVEC IL-11	17.2
Secondary Tr1 rest	0.0	Lung Microvascular EC none	79.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	11.2
Primary Th2 act	0.9	Microvascular Dermal EC none	1.9
Primary Tr1 act	1.4	Microvascular Dermal EC TNFalpha + IL-1beta	3.6

Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	26.4
Primary Th2 rest	0.3	Small airway epithelium none	15.7
Primary Tr1 rest	0.1	Small airway epithelium TNFalpha + IL-1beta	42.3
CD45RA CD4 lymphocyte act	16.8	Coronary artery SMC rest	27.0
CD45RO CD4 lymphocyte act	0.8	Coronary artery SMC TNFalpha + IL-1beta	45.4
CD8 lymphocyte act	0.0	Astrocytes rest	0.3
Secondary CD8 lymphocyte rest	0.8	Astrocytes TNFalpha + IL-1beta	2.0
Secondary CD8 lymphocyte act	0.4	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	1.5
2ry Th1/Th2/Tr1 _anti-CD95 CH11	0.1	CCD1106 (Keratinocytes) none	48.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	35.1
LAK cells IL-2	0.4	Liver cirrhosis	2.4
LAK cells IL-2+IL-12	0.0	NCI-H292 none	22.2
LAK cells IL-2+IFN gamma	0.6	NCI-H292 IL-4	16.0
LAK cells IL-2+ IL-18	0.2	NCI-H292 IL-9	30.8
LAK cells PMA/ionomycin	3.7	NCI-H292 IL-13	20.4
NK Cells IL-2 rest	1.9	NCI-H292 IFN gamma	13.5
Two Way MLR 3 day	0.1	HPAEC none	12.0
Two Way MLR 5 day	0.1	HPAEC TNF alpha + IL-1 beta	64.2
Two Way MLR 7 day	0.7	Lung fibroblast none	27.2
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	55.5
PBMC PWM	0.2	Lung fibroblast IL-4	35.8
PBMC PHA-L	0.1	Lung fibroblast IL-9	42.6
Ramos (B cell) none	0.0	Lung fibroblast IL-13	2.6
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	100.0
B lymphocytes PWM	0.8	Dermal fibroblast CCD1070 rest	31.2
B lymphocytes CD40L and IL-4	0.2	Dermal fibroblast CCD1070 TNF alpha	40.3
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	26.1
EOL-1 dbcAMP PMA/ionomycin	2.8	Dermal fibroblast IFN gamma	3.8
Dendritic cells none	0.0	Dermal fibroblast IL-4	1.8
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	2.8
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.8	Colon	0.5
Macrophages rest	0.0	Lung	2.2
Macrophages LPS	0.1	Thymus	0.6
HUVEC none	18.0	Kidney	7.1
HUVEC starved	14.2		

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag4510 Highest expression of this gene is detected in orthoarthritis bone (CT=29.5). This gene shows a widespread expression in this panel. Moderate to low levels of expression of this gene are detected in samples derived from normal and orthoarthritis/ rheumatoid arthritis bone, cartilage, synovium and  
5 synovial fluid samples, from normal lung, COPD lung, emphysema, atopic asthma, asthma, allergy, Crohn's disease (normal matched control and diseased), ulcerative colitis(normal matched control and diseased), and psoriasis (normal matched control and diseased). Therefore, therapeutic modulation of this gene product may ameliorate symptoms/conditions associated with autoimmune and inflammatory disorders including  
10 psoriasis, allergy, asthma, inflammatory bowel disease, rheumatoid arthritis and osteoarthritis

**General\_screening\_panel\_v1.4 Summary:** Ag4510 Highest expression of this gene is detected in a CNS cancer SF-295 cell line (CT=25.6). Moderate to high levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic,  
15 gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Expression of this gene is higher in cancer cell lines compared to the normal tissues. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung,  
20 liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, fetal skeletal muscle, heart, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of  
25 the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene  
30 product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Interestingly, this gene is expressed at much higher levels in fetal (CT=31-33) when compared to adult liver and skeletal muscle (CTs=35-38). This observation suggests that expression of this gene can be used to distinguish fetal from adult liver and skeletal muscle. In addition, the relative overexpression of this gene in fetal tissues suggest that the protein product may enhance liver and muscle growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of liver and muscle related diseases.

**Panel 4.1D Summary:** Ag4510 Highest expression of this gene is detected in a IFN gamma stimulated lung fibroblasts (CT=28.4). Moderate to low levels of expression of this gene is detected in endothelial cells, keratinocytes, dermal fibroblasts and lung related samples including resting and activated-NCI-H292 mucoepidermoid cells, resting and activated lung fibroblasts, human pulmonary aortic endothelial cells (treated and untreated), small airway epithelium (treated and untreated), treated bronchial epithelium and lung microvascular endothelial cells (treated and untreated). Low expression of this gene is also detected in activated secondary Th1, Th2 and Tr1 cells, activated eosinophils and activated CD45RA CD4 lymphocyte (CT=30.9), which represent activated naive T cells. In activated memory T cells (CD45RO CD4 lymphocyte) or CD4 Th1 or Th2 cells, resting CD4 cells (CTs>35), the expression of this gene is strongly down regulated suggesting a role for this putative protein in differentiation or activation of naive T cells. Therefore, therapeutic modulation of this gene may be useful in the treatment of autoimmune and inflammatory disorders that include arthritis, psoriasis, Crohns disease, ulcerative colitis, asthma, chronic obstructive pulmonary disease, allergy and emphysema.

**G. CG165719-01 (NOV12d), CG165719-02 (NOV12b) and CG165719-03 (NOV12c): NEURONAL MEMBRANE GLYCOPROTEIN M6-B.**

Expression of gene CG165719-01, CG165719-02 and CG165719-03 was assessed using the primer-probe sets Ag5977, Ag5978, Ag7810 and Ag7794, described in Tables GA, GB, GC and GD. Results of the RTQ-PCR runs are shown in Tables GE, GF and GG. Please note that primer-probe set Ag5977 is specific for CG165719-03 and Ag5978 is specific for CG165719-01 and CG165719-02.

**Table GA. Probe Name Ag5977**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - caagagagaaaaggctgctttg - 3'	22	109	132
Probe	TET-5' - ggaggagtcccctacgcctccct - 3' - TAMRA	23	151	133
Reverse	5' - cacagccgcagaataaggc - 3'	19	205	134

**Table GB. Probe Name Ag5978**

5

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - gtgaacagcagagctgaaatg - 3'	21	121	135
Probe	TET-5' - cccgtgcccaacctgggggacag - 3' - TAMRA	23	196	136
Reverse	5' - ggggactcctccagac - 3'	17	266	137

**Table GC. Probe Name Ag7810**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - gcatcagtggaaatgttcgttt - 3'	21	572	138
Probe	TET-5' - cagccaggccactccaagcacat - 3' - TAMRA	23	602	139
Reverse	5' - caccgctgagaaaccaaac - 3'	19	630	140

**Table GD. Probe Name Ag7794**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - gcgattcttgagcaaacactt - 3'	20	370	141
Probe	TET-5' - cacctcgctcagcaaggcatggt - 3' - TAMRA	23	407	142
Reverse	5' - ccatagatgacatactgcatcagtt - 3'	25	434	143

**Table GE. CNS neurodegeneration v1.0**

Tissue Name	Rel. Exp.(%) Ag5977, Run 248589057	Rel. Exp.(%) A5978, Run 248589058	Rel. Exp.(%) Ag7794, Run 312372407	Tissue Name	Rel. Exp.(%) Ag5977, Run 248589057	Rel. Exp.(%) Ag5978, Run 248589058	Rel. Exp.(%) Ag7794, Run 312372407
AD 1 Hippo	16.3	8.5	15.0	Control (Path) 3 Temporal Ctx	4.5	2.6	4.5

AD 2 Hippo	40.9	28.7	23.3	Control (Path) 4 Temporal Ctx	26.2	24.5	12.6
AD 3 Hippo	5.5	4.7	5.0	AD 1 Occipital Ctx	9.6	3.0	10.7
AD 4 Hippo	10.3	8.4	10.0	AD 2 Occipital Ctx (Missing)	0.0	0.0	0.0
AD 5 hippo	28.3	65.1	10.8	AD 3 Occipital Ctx	4.2	1.5	5.0
AD 6 Hippo	100.0	26.6	74.7	AD 4 Occipital Ctx	19.2	21.6	12.6
Control 2 Hippo	42.0	38.2	22.2	AD 5 Occipital Ctx	42.0	13.2	12.1
Control 4 Hippo	15.4	17.2	13.4	AD 6 Occipital Ctx	31.4	46.7	19.1
Control (Path) 3 Hippo	6.2	4.0	4.8	Control 1 Occipital Ctx	2.2	1.7	2.1
AD 1 Temporal Ctx	11.8	5.4	15.3	Control 2 Occipital Ctx	56.3	84.1	32.8
AD 2 Temporal Ctx	44.8	34.2	23.0	Control 3 Occipital Ctx	15.0	11.0	9.4
AD 3 Temporal Ctx	3.6	3.8	0.0	Control 4 Occipital Ctx	9.9	9.3	6.2
AD 4 Temporal Ctx	26.2	22.1	15.5	Control (Path) 1 Occipital Ctx	0.1	100.0	42.6
AD 5 Inf Temporal Ctx	39.0	82.9	100.0	Control (Path) 2 Occipital Ctx	7.4	6.4	5.5
AD 5 Sup Temporal Ctx	24.8	45.7	56.3	Control (Path) 3 Occipital Ctx	2.4	1.2	2.7
AD 6 Inf Temporal Ctx	74.7	47.6	45.1	Control (Path) 4 Occipital Ctx	5.6	4.9	4.7
AD 6 Sup Temporal Ctx	69.3	36.6	38.7	Control 1 Parietal Ctx	5.5	5.6	7.1
Control 1 Temporal Ctx	9.3	8.7	6.0	Control 2 Parietal Ctx	18.3	26.8	31.0
Control 2 Temporal Ctx	58.2	52.9	23.8	Control 3 Parietal Ctx	18.2	17.1	10.2

Control 3 Temporal Ctx	20.7	14.8	9.0	Control (Path) 1 Parietal Ctx	0.1	92.7	36.9
Control 4 Temporal Ctx	17.8	14.8	9.1	Control (Path) 2 Parietal Ctx	17.7	18.4	11.4
Control (Path) 1 Temporal Ctx	0.2	75.3	24.0	Control (Path) 3 Parietal Ctx	2.7	1.2	3.2
Control (Path) 2 Temporal Ctx	30.6	29.9	17.2	Control (Path) 4 Parietal Ctx	24.1	22.1	12.4

**Table GF. General screening panel v1.5**

Tissue Name	Rel. Exp.(%) Ag5977, Run 248220118	Rel. Exp.(%) g5978, Run 248445832	Tissue Name	Rel. Exp.(%) Ag5977, Run 248220118	Rel. Exp.(%) Ag5978, Run 248445832
Adipose	0.3	0.1	Renal ca. TK-10	0.0	0.0
Melanoma* Hs688(A).T	0.0	0.0	Bladder	0.6	0.7
Melanoma* Hs688(B).T	0.0	0.0	Gastric ca. (liver met.) NCI-N87	0.1	0.0
Melanoma* M14	0.7	3.4	Gastric ca. KATO III	0.0	0.0
Melanoma* LOXIMVI	0.0	0.0	Colon ca. SW-948	0.0	0.0
Melanoma* SK-MEL-5	2.1	11.7	Colon ca. SW480	0.0	0.0
Squamous cell carcinoma SCC-4	0.0	0.0	Colon ca.* (SW480 met) SW620	0.0	0.0
Testis Pool	0.5	0.8	Colon ca. HT29	0.0	0.0
Prostate ca.* (bone met) PC-3	0.0	0.0	Colon ca. HCT-116	0.0	0.0
Prostate Pool	2.8	2.0	Colon ca. CaCo-2	0.0	0.0
Placenta	0.0	0.0	Colon cancer tissue	0.0	0.0
Uterus Pool	2.3	1.0	Colon ca. SW1116	0.0	0.0
Ovarian ca. OVCAR-3	0.1	0.2	Colon ca. Colo-205	0.0	0.0
Ovarian ca. SK-OV-3	0.0	0.1	Colon ca. SW-48	0.0	0.0

Ovarian ca. OVCAR-4	0.0	0.0	Colon Pool	0.3	0.1
Ovarian ca. OVCAR-5	0.1	0.0	Small Intestine Pool	1.9	2.6
Ovarian ca. IGROV-1	54.0	24.7	Stomach Pool	1.3	1.1
Ovarian ca. OVCAR-8	3.9	5.9	Bone Marrow Pool	0.8	1.2
Ovary	0.0	0.0	Fetal Heart	0.7	0.5
Breast ca. MCF-7	0.0	0.0	Heart Pool	0.7	1.1
Breast ca. MDA-MB-23 1	0.0	0.0	Lymph Node Pool	0.4	0.9
Breast ca. BT 549	0.0	0.0	Fetal Skeletal Muscle	0.3	0.5
Breast ca. T47D	0.0	0.0	Skeletal Muscle Pool	1.1	1.0
Breast ca. MDA-N	0.2	1.8	Spleen Pool	0.1	0.7
Breast Pool	0.1	0.1	Thymus Pool	0.1	0.5
Trachea	2.2	0.6	CNS cancer (glio/astro) U87-MG	0.0	0.0
Lung	0.3	0.3	CNS cancer (glio/astro) U-118-MG	0.0	0.0
Fetal Lung	0.8	1.2	CNS cancer (neuro;met) SK-N-AS	0.0	0.3
Lung ca. NCI-N417	0.0	0.3	CNS cancer (astro) SF-539	0.0	0.0
Lung ca. LX-1	0.0	0.0	CNS cancer (astro) SNB-75	62.9	46.3
Lung ca. NCI-H146	0.2	1.7	CNS cancer (glio) SNB-19	73.2	27.9
Lung ca. SHP-77	0.0	0.0	CNS cancer (glio) SF-295	0.0	0.0
Lung ca. A549	0.0	0.0	Brain (Amygdala) Pool	84.1	43.2
Lung ca. NCI-H526	0.1	0.0	Brain (cerebellum)	98.6	100.0
Lung ca. NCI-H23	0.0	0.1	Brain (fetal)	53.6	24.8
Lung ca. NCI-H460	0.0	0.0	Brain (Hippocampus) Pool	100.0	45.4
Lung ca. HOP-62	0.0	0.0	Cerebral Cortex Pool	89.5	54.0

Lung ca. NCI-H522	0.0	0.0	Brain (Substantia nigra) Pool	94.6	41.5
Liver	0.0	0.0	Brain (Thalamus) Pool	87.1	70.2
Fetal Liver	0.0	0.0	Brain (whole)	57.4	48.6
Liver ca. HepG2	0.0	0.0	Spinal Cord Pool	57.4	27.0
Kidney Pool	0.4	1.1	Adrenal Gland	0.4	0.4
Fetal Kidney	0.1	0.1	Pituitary gland Pool	1.3	1.2
Renal ca. 786-0	0.0	0.0	Salivary Gland	0.4	0.8
Renal ca. A498	0.0	0.0	Thyroid (female)	0.3	0.1
Renal ca. ACHN	0.0	0.0	Pancreatic ca. CAPAN2	0.0	0.0
Renal ca. UO-31	0.0	0.0	Pancreas Pool	0.4	0.6

**Table GG. Panel 4.1D**

Tissue Name	Rel. Exp.(%) Ag578, Run 2481226 33	Rel. Exp.(%) Ag7794, Run 3123559 78	Rel. Exp.(%) Ag7810, Run 3123633 84	Tissue Name	Rel. Exp.(%) Ag5978, Run 2481226 33	Rel. Exp.(%) Ag7794, Run 3123559 78	Rel. Exp.(%) Ag7810, Run 3123633 84
Secondary Th1 act	0.0	0.0	0.0	HUVEC IL-1beta	0.0	0.0	0.0
Secondary Th2 act	0.0	0.0	0.0	HUVEC IFN gamma	0.0	0.0	0.0
Secondary Tr1 act	0.0	0.0	0.0	HUVEC TNF alpha + IFN gamma	0.0	0.0	0.0
Secondary Th1 rest	0.0	0.0	0.0	HUVEC TNF alpha + IL4	0.0	0.0	0.0
Secondary Th2 rest	0.0	0.0	0.0	HUVEC IL-11	0.0	1.4	0.0
Secondary Tr1 rest	0.0	0.0	0.0	Lung Microvascular EC none	0.0	0.0	0.0
Primary Th1 act	0.0	0.0	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0	0.0	0.0
Primary Th2 act	0.0	0.0	0.0	Microvascular Dermal EC none	0.0	0.0	0.0
Primary Tr1 act	0.0	0.0	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0	0.4	0.0
Primary Th1 rest	0.0	0.0	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0	0.0	0.0
Primary Th2 rest	0.0	0.0	0.0	Small airway epithelium none	0.0	1.7	4.9
Primary Tr1 rest	0.0	0.0	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0	0.0	0.0
CD45RA CD4 lymphocyte act	0.0	0.0	0.0	Coronary artery SMC rest	0.0	3.0	0.0

CD45RO CD4 lymphocyte act	0.0	0.0	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0	2.3
CD8 lymphocyte act	0.0	0.0	0.0	Astrocytes rest	100.0	100.0	100.0
Secondary CD8 lymphocyte rest	0.0	0.0	0.0	Astrocytes TNFalpha + IL-1beta	15.3	14.7	34.2
Secondary CD8 lymphocyte act	0.0	0.0	0.0	KU-812 (Basophil) rest	0.0	0.0	0.0
CD4 lymphocyte none	0.0	0.8	0.0	KU-812 (Basophil) PMA/ionomycin	0.0	0.0	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	0.0	0.0	CCD1106 (Keratinocytes) none	0.0	11.9	40.3
LAK cells rest	0.0	0.0	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0	1.0	1.1
LAK cells IL-2	0.0	0.0	0.0	Liver cirrhosis	0.0	4.5	8.8
LAK cells IL-2+IL-12	0.0	0.0	0.0	NCI-H292 none	0.0	0.0	0.0
LAK cells IL-2+IFN gamma	0.0	0.0	0.0	NCI-H292 IL-4	0.0	0.0	0.0
LAK cells IL-2+IL-18	0.0	0.0	0.0	NCI-H292 IL-9	0.0	3.0	0.0
LAK cells PMA/ionomycin	0.0	0.0	0.0	NCI-H292 IL-13	0.0	0.0	5.5
NK Cells IL-2 rest	0.0	0.0	2.3	NCI-H292 IFN gamma	0.0	0.0	0.0
Two Way MLR 3 day	0.0	0.0	0.0	HPAEC none	0.0	0.0	0.0
Two Way MLR 5 day	0.0	0.0	0.0	HPAEC TNF alpha + IL-1 beta	0.0	2.3	7.9
Two Way MLR 7 day	0.0	0.8	0.0	Lung fibroblast none	0.0	11.6	35.8
PBMC rest	0.0	0.0	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0	7.1	7.8
PBMC PWM	0.0	0.0	0.0	Lung fibroblast IL-4	0.0	5.7	3.0
PBMC PHA-L	0.0	0.0	0.0	Lung fibroblast IL-9	0.0	3.6	11.4

Ramos (B cell) none	0.0	0.0	0.0	Lung fibroblast IL-13	0.0	4.5	5.4
Ramos (B cell) ionomycin	0.0	0.0	0.0	Lung fibroblast IFN gamma	27.0	19.1	58.2
B lymphocytes PWM	0.0	0.0	0.0	Dermal fibroblast CCD1070 rest	0.0	8.1	1.6
B lymphocytes CD40L and IL-4	0.0	0.0	0.0	Dermal fibroblast CCD1070 TNF alpha	29.1	4.0	1.5
EOL-1 dbcAMP	0.0	0.0	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0	2.9	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	0.0	0.0	Dermal fibroblast IFN gamma	0.0	5.2	19.9
Dendritic cells none	0.0	0.0	0.0	Dermal fibroblast IL-4	0.0	8.5	32.3
Dendritic cells LPS	0.0	0.0	0.0	Dermal Fibroblasts rest	0.0	0.9	2.6
Dendritic cells anti-CD40	0.0	0.0	0.0	Neutrophils TNFa+LPS	0.0	0.0	0.0
Monocytes rest	0.0	0.0	0.0	Neutrophils rest	0.0	0.0	0.0
Monocytes LPS	0.0	0.0	0.0	Colon	16.0	10.1	1.1
Macrophages rest	0.0	0.0	0.0	Lung	0.0	6.0	8.0
Macrophages LPS	0.0	0.0	0.0	Thymus	0.0	0.0	0.0
HUVEC none	0.0	0.0	0.0	Kidney	0.0	10.3	64.6
HUVEC starved	0.0	0.0	0.0				

- CNS\_neurodegeneration\_v1.0 Summary:** Ag5977/Ag5978/Ag7794 Three experiments with different probe pimer sets are in good agreement. This panel confirms the expression of this gene at significant levels in the brains of an independent group of
- 5 individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.5 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**General\_screening\_panel\_v1.5 Summary:** Ag5977/Ag5978 Two experiments with different probe primer sets are in good agreement with highest expression of this gene seen in cerebellum and hippocampus (CTs=27-28.9). This gene shows preferential expression in all the regions of brain including including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Moderate expression of this gene is also seen in two of the brain cancer, two ovarian cancer and melanoma cell lines. Therefore, therapeutic modulation of this gene may be useful in the treatment of melanoma, brain, and ovarian cancers.

Low levels of expression of this gene is also seen in pancreas, pituitary gland, skeletal muscle and gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

**Panel 4.1D Summary:** Ag5978/Ag7794/Ag7810 Multiple experiments with different probe-primer sets are in good agreement. Highest expression of this gene is detected in resting astrocytes (CTs=31-34.7). Low expression of this gene is also seen in activated astrocytes and lung fibroblasts. Therefore, therapeutic regulation of this gene or the encoded protein could be important in the treatment of multiple sclerosis or other inflammatory diseases of the CNS and and inflammatory lung disorders including chronic obstructive pulmonary disease, asthma, allergy and emphysema.

Ag5977 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

#### **H. CG167488-01 (NOV13b): Hypothetical Transmembrane Protein.**

Expression of gene CG167488-01 was assessed using the primer-probe set Ag5997, described in Table HA. Results of the RTQ-PCR runs are shown in Tables HB, HC and HD.

**Table HA. Probe Name Ag5997**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gagctaccttataaagaccatctgtacat-3'	29	3	144
Probe	TET-5'-ccactgtgaaatggagtttcaaaatcaca-3'-TAMR A	29	32	145
Reverse	5'-atatgtgctcctagtccttatgttcatgt-3'	28	73	146

**Table HB. CNS neurodegeneration v1.0**

5

Tissue Name	Rel. Exp.(%) Ag5997 Run 248589037	Tissue Name	Rel. Exp.(%) Ag5997, Run 248589037
AD 1 Hippo	0.0	Control (Path) 3 Temporal Ctx	2.6
AD 2 Hippo	28.9	Control (Path) 4 Temporal Ctx	54.3
AD 3 Hippo	1.6	AD 1 Occipital Ctx	4.6
AD 4 Hippo	16.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	92.7	AD 3 Occipital Ctx	0.0
AD 6 Hippo	31.4	AD 4 Occipital Ctx	36.1
Control 2 Hippo	35.4	AD 5 Occipital Ctx	48.0
Control 4 Hippo	16.8	AD 6 Occipital Ctx	41.5
Control (Path) 3 Hippo	28.9	Control 1 Occipital Ctx	0.0
AD 1 Temporal Ctx	4.6	Control 2 Occipital Ctx	45.4
AD 2 Temporal Ctx	35.4	Control 3 Occipital Ctx	20.0
AD 3 Temporal Ctx	9.1	Control 4 Occipital Ctx	3.3
AD 4 Temporal Ctx	30.1	Control (Path) 1 Occipital Ctx	99.3
AD 5 Inf Temporal Ctx	83.5	Control (Path) 2 Occipital Ctx	12.9
AD 5 SupTemporal Ctx	68.8	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	88.3	Control (Path) 4 Occipital Ctx	31.2
AD 6 Sup Temporal Ctx	59.5	Control 1 Parietal Ctx	1.5
Control 1 Temporal Ctx	1.2	Control 2 Parietal Ctx	47.6
Control 2 Temporal Ctx	34.2	Control 3 Parietal Ctx	20.6
Control 3 Temporal Ctx	36.3	Control (Path) 1 Parietal Ctx	100.0
Control 4 Temporal Ctx	13.8	Control (Path) 2 Parietal Ctx	29.9
Control (Path) 1 Temporal Ctx	90.8	Control (Path) 3 Parietal Ctx	0.0
Control (Path) 2 Temporal Ctx	52.9	Control (Path) 4 Parietal Ctx	75.3

Table HC. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag5997, Run 248592793	issue Name	Rel. Exp.(%) Ag5997, Run 248592793
Adipose	36.3	Renal ca. TK-10	9.2
Melanoma* Hs688(A).T	2.9	Bladder	14.2
Melanoma* Hs688(B).T	7.2	Gastric ca. (liver met.) NCI-N87	22.4
Melanoma* M14	0.0	Gastric ca. KATO III	12.5
Melanoma* LOXIMVI	3.4	Colon ca. SW-948	6.2
Melanoma* SK-MEL-5	3.5	Colon ca. SW480	10.7
Squamous cell carcinoma SCC-4	5.4	Colon ca.* (SW480 met) SW620	1.5
Testis Pool	8.2	Colon ca. HT29	7.3
Prostate ca.* (bone met) PC-3	5.0	Colon ca. HCT-116	12.3
Prostate Pool	13.2	Colon ca. CaCo-2	4.2
Placenta	8.8	Colon cancer tissue	11.8
Uterus Pool	5.9	Colon ca. SW1116	3.0
Ovarian ca. OVCAR-3	12.6	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	22.1	Colon ca. SW-48	0.4
Ovarian ca. OVCAR-4	3.1	Colon Pool	12.2
Ovarian ca. OVCAR-5	25.9	Small Intestine Pool	5.2
Ovarian ca. IGROV-1	6.0	Stomach Pool	13.4
Ovarian ca. OVCAR-8	7.2	Bone Marrow Pool	3.7
Ovary	1.6	Fetal Heart	0.8
Breast ca. MCF-7	0.6	Heart Pool	7.0
Breast ca. MDA-MB-231	30.6	Lymph Node Pool	9.7
Breast ca. BT 549	7.9	Fetal Skeletal Muscle	3.2
Breast ca. T47D	0.0	Skeletal Muscle Pool	18.0
Breast ca. MDA-N	0.1	Spleen Pool	1.3
Breast Pool	13.8	Thymus Pool	6.7
Trachea	12.1	CNS cancer (glio/astro) U87-MG	12.8
Lung	0.5	CNS cancer (glio/astro) U-118-MG	14.8
Fetal Lung	100.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	6.8
Lung ca. LX-1	3.4	CNS cancer (astro) SNB-75	7.7
Lung ca. NCI-H146	4.1	CNS cancer (glio) SNB-19	5.1
Lung ca. SHP-77	12.4	CNS cancer (glio) SF-295	14.9
Lung ca. A549	10.3	Brain (Amygdala) Pool	5.4
Lung ca. NCI-H526	0.0	Brain (cerebellum)	3.5
Lung ca. NCI-H23	6.0	Brain (fetal)	0.1
Lung ca. NCI-H460	7.6	Brain (Hippocampus) Pool	6.8
Lung ca. HOP-62	3.9	Cerebral Cortex Pool	7.7
Lung ca. NCI-H522	11.5	Brain (Substantia nigra) Pool	4.0

Liver	0.0	Brain (Thalamus) Pool	10.7
Fetal Liver	1.8	Brain (whole)	6.3
Liver ca. HepG2	1.3	Spinal Cord Pool	5.9
Kidney Pool	10.8	Adrenal Gland	0.1
Fetal Kidney	5.6	Pituitary gland Pool	5.8
Renal ca. 786-0	10.9	Salivary Gland	2.3
Renal ca. A498	3.8	Thyroid (female)	6.6
Renal ca. ACHN	4.3	Pancreatic ca. CAPAN2	21.8
Renal ca. UO-31	7.9	Pancreas Pool	17.2

**Table HD. Panel 5D**

	Rel. Exp.(%) Ag5997, Run 263248222	Tissue Name	Rel. Exp.(%) Ag5997, Run 263248222
97457_Patient-02go_adipose	13.4	94709_Donor 2 AM - A_adipose	0.0
97476_Patient-07sk_skeletal muscle	0.0	94710_Donor 2 AM - B_adipose	9.5
97477_Patient-07ut_uterus	0.0	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	22.8	94712_Donor 2 AD - A_adipose	0.0
97481_Patient-08sk_skeletal muscle	1.6	94713_Donor 2 AD - B_adipose	7.2
97482_Patient-08ut_uterus	0.0	94714_Donor 2 AD - C_adipose	0.0
97483_Patient-08pl_placenta	11.2	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.0
97486_Patient-09sk_skeletal muscle	0.0	94743_Donor 3 U - B_Mesenchymal Stem Cells	5.6
97487_Patient-09ut_uterus	13.8	94730_Donor 3 AM - A_adipose	4.5
97488_Patient-09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	0.0
97492_Patient-10ut_uterus	29.5	94732_Donor 3 AM - C_adipose	3.2
97493_Patient-10pl_placenta	0.0	94733_Donor 3 AD - A_adipose	0.0
97495_Patient-11go_adipose	17.7	94734_Donor 3 AD - B_adipose	0.0
97496_Patient-11sk_skeletal muscle	0.0	94735_Donor 3 AD - C_adipose	9.2
97497_Patient-11ut_uterus	58.6	77138_Liver_HepG2untreated	8.2
97498_Patient-11pl_placenta	100.0	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	0.0	81735_Small Intestine	5.4
97501_Patient-12sk_skeletal muscle	0.0	72409_Kidney_Proximal Convoluted Tubule	2.5
97502_Patient-12ut_uterus	5.9	82685_Small intestine_Duodenum	0.0
97503_Patient-12pl_placenta	6.2	90650_Adrenal_Adrenocortical adenoma	0.0

94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	21.2
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	5.8
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5997 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's  
5 diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.5 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**General\_screening\_panel\_v1.5 Summary:** Ag5997 Highest expression of this gene is detected in fetal lung (CT=29.4). Interestingly, this gene is expressed at much  
10 higher levels in fetal compared to adult lung (CT=37). This observation suggests that expression of this gene can be used to distinguish fetal from adult lung. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance lung growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene  
15 could be useful in treatment of lung related diseases.

Moderate to low levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic  
20 modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adipose, thyroid, pituitary gland, skeletal muscle, heart, and  
25 the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at low levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as

- 5 Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

**Panel 5D Summary:** Ag5997 Low expression of this gene is exclusively seen in placenta of non-diabetic but obese patient (CT=33.9). Therefore, expression of this gene may be used to distinguish placenta from other samples used in this panel.

- 10 **I. CG50970-01 (NOV15b) and CG50970-02 (NOV15i): Glypican-2 precursor.**

Expression of gene CG50970-01 and CG50970-03 was assessed using the primer-probe sets Ag1309 and Ag2251, described in Tables IA and IB. Results of the RTQ-PCR runs are shown in Tables IC, ID, IE, IF, IG, IH, II, IJ and IK. Please note that

15 CG50970-03 represents a full-length physical clone.

**Table IA. Probe Name Ag1309**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-actctctgacccagctcttctc-3'	22	359	147
Probe	TET-5'-ccactcctacggccgcctgtatg-3'-TAMRA	23	381	148
Reverse	5'-gagaacaggccattgaatatga-3'	22	416	149

**Table IB. Probe Name Ag2251**

20

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-actctctgacccagctcttctc-3'	22	359	150
Probe	TET-5'-ccactcctacggccgcctgtatg-3'-TAMRA	23	381	151
Reverse	5'-gagaacaggccattgaatatga-3'	22	416	152

**Table IC. AI comprehensive panel v1.0**

Tissue Name	Rel. Exp.(%) Ag2251, Run 44570248	Tissue Name	Rel. Exp.(%) Ag2251, Run 244570248
110967 COPD-F	20.7	112427 Match Control Psoriasis-F	29.3
110980 COPD-F	6.0	112418 Psoriasis-M	4.7
110968 COPD-M	4.4	112723 Match Control Psoriasis-M	27.9
110977 COPD-M	8.8	112419 Psoriasis-M	2.1
110989 Emphysema-F	12.6	112424 Match Control Psoriasis-M	2.9
110992 Emphysema-F	2.9	112420 Psoriasis-M	20.4
110993 Emphysema-F	16.4	112425 Match Control Psoriasis-M	23.8
110994 Emphysema-F	3.8	104689 (MF) OA Bone-Backus	8.7
110995 Emphysema-F	19.3	104690 (MF) Adj "Normal" Bone-Backus	8.4
110996 Emphysema-F	2.4	104691 (MF) OA Synovium-Backus	4.3
110997 Asthma-M	5.6	104692 (BA) OA Cartilage-Backus	2.4
111001 Asthma-F	14.8	104694 (BA) OA Bone-Backus	4.6
111002 Asthma-F	16.4	104695 (BA) Adj "Normal" Bone-Backus	7.7
111003 Atopic Asthma-F	16.2	104696 (BA) OA Synovium-Backus	2.7
111004 Atopic Asthma-F	28.3	104700 (SS) OA Bone-Backus	9.0
111005 Atopic Asthma-F	7.2	104701 (SS) Adj "Normal" Bone-Backus	3.8
111006 Atopic Asthma-F	4.4	104702 (SS) OA Synovium-Backus	7.5
111417 Allergy-M	11.0	117093 OA Cartilage Rep7	14.7
112347 Allergy-M	7.5	112672 OA Bone5	57.0
112349 Normal Lung-F	9.4	112673 OA Synovium5	27.9
112357 Normal Lung-F	34.2	112674 OA Synovial Fluid cells5	24.8
112354 Normal Lung-M	9.2	117100 OA Cartilage Rep14	4.0
112374 Crohns-F	10.3	112756 OA Bone9	100.0
112389 Match Control Crohns-F	6.0	112757 OA Synovium9	17.9
112375 Crohns-F	22.2	112758 OA Synovial Fluid Cells9	9.2
112732 Match Control Crohns-F	7.5	117125 RA Cartilage Rep2	6.9
112725 Crohns-M	0.0	113492 Bone2 RA	3.7
112387 Match Control Crohns-M	3.0	113493 Synovium2 RA	0.6
112378 Crohns-M	10.4	113494 Syn Fluid Cells RA	3.0
112390 Match Control Crohns-M	40.6	113499 Cartilage4 RA	1.3
112726 Crohns-M	6.7	113500 Bone4 RA	2.2
112731 Match Control Crohns-M	9.0	113501 Synovium4 RA	2.8
112380 Ulcer Col-F	25.5	113502 Syn Fluid Cells4 RA	5.3
112734 Match Control Ulcer Col-F	9.5	113495 Cartilage3 RA	0.0
112384 Ulcer Col-F	22.2	113496 Bone3 RA	2.9

112737 Match Control Ulcer Col-F	5.3	113497 Synovium3 RA	0.0
112386 Ulcer Col-F	0.0	113498 Syn Fluid Cells3 RA	0.0
112738 Match Control Ulcer Col-F	0.0	117106 Normal Cartilage Rep20	0.0
112381 Ulcer Col-M	2.0	113663 Bone3 Normal	10.3
112735 Match Control Ulcer Col-M	6.9	113664 Synovium3 Normal	6.5
112382 Ulcer Col-M	15.8	113665 Syn Fluid Cells3 Normal	3.6
112394 Match Control Ulcer Col-M	5.6	117107 Normal Cartilage Rep22	10.5
112383 Ulcer Col-M	13.3	113667 Bone4 Normal	9.3
112736 Match Control Ulcer Col-M	2.3	113668 Synovium4 Normal	22.7
112423 Psoriasis-F	4.1	113669 Syn Fluid Cells4 Normal	12.2

**Table ID. CNS neurodegeneration v1.0**

Tissue Name	Rel. Exp.(%) Ag225, Run 206265375	Tissue Name	Rel. Exp.(%) Ag2251, Run 206265375
AD 1 Hippo	19.8	Control (Path) 3 Temporal Ctx	3.5
AD 2 Hippo	35.8	Control (Path) 4 Temporal Ctx	34.2
AD 3 Hippo	7.2	AD 1 Occipital Ctx	19.5
AD 4 Hippo	9.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	59.0	AD 3 Occipital Ctx	17.3
AD 6 Hippo	97.9	AD 4 Occipital Ctx	19.2
Control 2 Hippo	37.1	AD 5 Occipital Ctx	52.9
Control 4 Hippo	34.9	AD 6 Occipital Ctx	42.0
Control (Path) 3 Hippo	17.2	Control 1 Occipital Ctx	6.3
AD 1 Temporal Ctx	20.6	Control 2 Occipital Ctx	51.8
AD 2 Temporal Ctx	33.7	Control 3 Occipital Ctx	23.0
AD 3 Temporal Ctx	9.9	Control 4 Occipital Ctx	6.6
AD 4 Temporal Ctx	36.1	Control (Path) 1 Occipital Ctx	73.7
AD 5 Inf Temporal Ctx	76.8	Control (Path) 2 Occipital Ctx	16.8
AD 5 Sup Temporal Ctx	97.9	Control (Path) 3 Occipital Ctx	11.8
AD 6 Inf Temporal Ctx	59.9	Control (Path) 4 Occipital Ctx	28.1
AD 6 Sup Temporal Ctx	100.0	Control 1 Parietal Ctx	12.1
Control 1 Temporal Ctx	9.9	Control 2 Parietal Ctx	62.4
Control 2 Temporal Ctx	29.9	Control 3 Parietal Ctx	20.4
Control 3 Temporal Ctx	10.5	Control (Path) 1 Parietal Ctx	43.8
Control 4 Temporal Ctx	34.2	Control (Path) 2 Parietal Ctx	14.9
Control (Path) 1 Temporal Ctx	63.7	Control (Path) 3 Parietal Ctx	7.8

Control (Path) 2 Temporal Ctx	13.8	Control (Path) 4 Parietal Ctx	29.9
-------------------------------	------	-------------------------------	------

**Table IE. General screening panel v1.5**

Tissue Name	Rel. Exp.(%) Ag2251, Run 246733742	issue Name	Rel. Exp.(%) Ag2251, Run 246733742
Adipose	0.2	Renal ca. TK-10	9.2
Melanoma* Hs688(A).T	0.8	Bladder	0.6
Melanoma* Hs688(B).T	0.9	Gastric ca. (liver met.) NCI-N87	0.3
Melanoma* M14	4.5	Gastric ca. KATO III	1.4
Melanoma* LOXIMVI	0.6	Colon ca. SW-948	0.1
Melanoma* SK-MEL-5	4.3	Colon ca. SW480	3.8
Squamous cell carcinoma SCC-4	0.4	Colon ca.* (SW480 met) SW620	1.6
Testis Pool	8.1	Colon ca. HT29	0.9
Prostate ca.* (bone met) PC-3	3.6	Colon ca. HCT-116	2.5
Prostate Pool	0.0	Colon ca. CaCo-2	4.3
Placenta	0.7	Colon cancer tissue	0.7
Uterus Pool	0.1	Colon ca. SW1116	0.7
Ovarian ca. OVCAR-3	2.9	Colon ca. Colo-205	0.2
Ovarian ca. SK-OV-3	0.5	Colon ca. SW-48	0.8
Ovarian ca. OVCAR-4	0.6	Colon Pool	0.8
Ovarian ca. OVCAR-5	1.2	Small Intestine Pool	0.9
Ovarian ca. IGROV-1	5.5	Stomach Pool	0.4
Ovarian ca. OVCAR-8	1.6	Bone Marrow Pool	0.3
Ovary	0.8	Fetal Heart	2.0
Breast ca. MCF-7	1.6	Heart Pool	0.3
Breast ca. MDA-MB-231	0.2	Lymph Node Pool	0.8
Breast ca. BT 549	22.5	Fetal Skeletal Muscle	2.8
Breast ca. T47D	0.2	Skeletal Muscle Pool	0.3
Breast ca. MDA-N	2.6	Spleen Pool	0.3
Breast Pool	1.3	Thymus Pool	9.9
Trachea	0.3	CNS cancer (glio/astro) U87-MG	3.0
Lung	0.4	CNS cancer (glio/astro) U-118-MG	0.8
Fetal Lung	2.9	CNS cancer (neuro;met) SK-N-AS	22.4
Lung ca. NCI-N417	5.0	CNS cancer (astro) SF-539	0.3
Lung ca. LX-1	5.3	CNS cancer (astro) SNB-75	17.2
Lung ca. NCI-H146	62.9	CNS cancer (glio) SNB-19	7.3
Lung ca. SHP-77	12.4	CNS cancer (glio) SF-295	7.6
Lung ca. A549	1.5	Brain (Amygdala) Pool	1.5
Lung ca. NCI-H526	25.9	Brain (cerebellum)	3.1
Lung ca. NCI-H23	7.0	Brain (fetal)	100.0

Lung ca. NCI-H460	6.7	Brain (Hippocampus) Pool	1.0
Lung ca. HOP-62	1.3	Cerebral Cortex Pool	1.3
Lung ca. NCI-H522	36.3	Brain (Substantia nigra) Pool	0.8
Liver	0.0	Brain (Thalamus) Pool	2.1
Fetal Liver	0.4	Brain (whole)	2.2
Liver ca. HepG2	0.2	Spinal Cord Pool	1.7
Kidney Pool	0.9	Adrenal Gland	0.3
Fetal Kidney	9.3	Pituitary gland Pool	0.1
Renal ca. 786-0	0.8	Salivary Gland	0.2
Renal ca. A498	0.7	Thyroid (female)	0.0
Renal ca. ACHN	1.0	Pancreatic ca. CAPAN2	0.8
Renal ca. UO-31	5.7	Pancreas Pool	0.9

**Table IF. Oncology cell line screening panel v3.2**

Tissue Name	Rel. Exp.(%) g2251, Run 2482021 32	Tissue Name	Rel. Exp.(%) Ag2251, Run 2482021 32
94905_Daoy_Medulloblastoma/Cerebellu m_sscDNA	1.2	94954_Ca Ski_Cervical epidermoid carcinoma (metastasis)_sscDNA	10.7
94906_TE671_Medulloblastom/Cerebellu m_sscDNA	14.8	94955_ES-2_Ovarian clear cell carcinoma_sscDNA	2.1
94907_D283 Med_Medulloblastoma/Cerebellum_sscDN A	16.6	94957_Ramos/6h stim_Stimulated with PMA/ionomycin 6h_sscDNA	4.0
94908_PFSK-1_Primitive Neuroectodermal/Cerebellum_sscDNA	1.7	94958_Ramos/14h stim_ Stimulated with PMA/ionomycin 14h_sscDNA	2.9
94909_XF-498_CNS_sscDNA	1.2	94962_MEG-01_Chronic myelogenous leukemia (megokaryoblast)_sscDNA	0.3
94910_SNB-78_CNS/glioma_sscDNA	1.0	94963_Raji_Burkitt's lymphoma_sscDNA	1.2
94911_SF-268_CNS/glioblastoma_sscDN A	1.8	94964_Daudi_Burkitt's lymphoma_sscDNA	2.4
94912_T98G_Glioblastoma_sscDNA	1.3	94965_U266_B-cell plasmacytoma/myeloma_sscDNA	0.0
96776_SK-N-SH_Neuroblastoma (metastasis)_sscDNA	11.7	94968_CA46_Burkitt's lymphoma_sscDNA	3.7
94913_SF-295_CNS/glioblastoma_sscDN A	1.4	94970_RL_non-Hodgkin's B-cell lymphoma_sscDNA	3.3
132565_NT2 pool_sscDNA	16.6	94972_JM1_pre-B-cell lymphoma/leukemia_sscDNA	5.5

94914_Cerebellum_sscDNA	1.3	94973_Jurkat_T cell leukemia_sscDNA	5.1
96777_Cerebellum_sscDNA	1.7	94974_TF-1_Erythroleukemia_sscDNA	1.5
94916_NCI-H292_Mucoepidermoid lung carcinoma_sscDNA	0.3	94975_HUT 78_T-cell lymphoma_sscDNA	2.5
94917_DMS-114_Small cell lung cancer_sscDNA	24.7	94977_U937_Histiocytic lymphoma_sscDNA	1.4
94918_DMS-79_Small cell lung cancer/neuroendocrine_sscDNA	100.0	94980_KU-812_Myelogenous leukemia_sscDNA	0.0
94919_NCI-H146_Small cell lung cancer/neuroendocrine_sscDNA	80.1	94981_769-P_Clear cell renal carcinoma_sscDNA	0.0
94920_NCI-H526_Small cell lung cancer/neuroendocrine_sscDNA	79.0	94983_Caki-2_Clear cell renal carcinoma_sscDNA	0.1
94921_NCI-N417_Small cell lung cancer/neuroendocrine_sscDNA	7.9	94984_SW 839_Clear cell renal carcinoma_sscDNA	0.9
94923_NCI-H82_Small cell lung cancer/neuroendocrine_sscDNA	23.2	94986_G401_Wilms' tumor_sscDNA	0.9
94924_NCI-H157_Squamous cell lung cancer (metastasis)_sscDNA	0.2	126768_293 cells_sscDNA	1.9
94925_NCI-H1155_Large cell lung cancer/neuroendocrine_sscDNA	19.5	94987_Hs766T_Pancreatic carcinoma (LN metastasis)_sscDNA	0.7
94926_NCI-H1299_Large cell lung cancer/neuroendocrine_sscDNA	9.3	94988_CAPAN-1_Pancreatic adenocarcinoma (liver metastasis)_sscDNA	0.1
94927_NCI-H727_Lung carcinoid_sscDNA	0.9	94989_SU86.86_Pancreatic carcinoma (liver metastasis)_sscDNA	0.0
94928_NCI-UMC-11_Lung carcinoid_sscDNA	5.1	94990_BxPC-3_Pancreatic adenocarcinoma_sscDNA	0.0
94929_LX-1_Small cell lung cancer_sscDNA	1.0	94991_HPAC_Pancreatic adenocarcinoma_sscDNA	0.6
94930_Colo-205_Colon cancer_sscDNA	0.0	94992_MIA PaCa-2_Pancreatic carcinoma_sscDNA	0.0
94931_KM12_Colon cancer_sscDNA	0.6	94993_CFPAC-1_Pancreatic ductal adenocarcinoma_sscDNA	1.0
94932_KM20L2_Colon cancer_sscDNA	0.2	94994_PANC-1_Pancreatic epithelioid ductal carcinoma_sscDNA	0.2
94933_NCI-H716_Colon cancer_sscDNA	5.6	94996_T24_Bladder carcinoma (transitional cell)_sscDNA	0.2
94935_SW-48_Colon adenocarcinoma_sscDNA	0.2	94997_5637_Bladder carcinoma_sscDNA	2.4
94936_SW1116_Colon adenocarcinoma_sscDNA	0.3	94998_HT-1197_Bladder carcinoma_sscDNA	0.7
94937_LS 174T_Colon adenocarcinoma_sscDNA	0.3	94999_UM-UC-3_Bladder carcinoma (transitional cell)_sscDNA	0.2

94938_SW-948_Colon adenocarcinoma_sscDNA	0.2	95000_A204_Rhabdomyosarcoma_sscDNA	0.0
94939_SW-480_Colon adenocarcinoma_sscDNA	0.0	95001_HT-1080_Fibrosarcoma_sscDNA	0.0
94940_NCI-SNU-5_Gastric carcinoma_sscDNA	0.9	95002_MG-63_Osteosarcoma (bone)_sscDNA	0.0
112197_KATO III_Stomach_sscDNA	0.0	95003_SK-LMS-1_Leiomyosarcoma (vulva)_sscDNA	1.3
94943_NCI-SNU-16_Gastric carcinoma_sscDNA	0.2	95004_SJRH30_Rhabdomyosarcoma (met to bone marrow)_sscDNA	9.9
94944_NCI-SNU-1_Gastric carcinoma_sscDNA	1.5	95005_A431_Epidermoid carcinoma_sscDNA	0.2
94946_RF-1_Gastric adenocarcinoma_sscDNA	0.3	95007_WM266-4_Melanoma_sscDNA	0.9
94947_RF-48_Gastric adenocarcinoma_sscDNA	0.6	112195_DU 145_Prostate_sscDNA	0.2
96778_MKN-45_Gastric carcinoma_sscDNA	0.6	95012_MDA-MB-468_Breast adenocarcinoma_sscDNA	0.5
94949_NCI-N87_Gastric carcinoma_sscDNA	0.6	112196_SSC-4_Tongue_sscDNA	0.9
94951_OVCAR-5_Ovarian carcinoma_sscDNA	0.0	112194_SSC-9_Tongue_sscDNA	1.3
94952_RL95-2_Uterine carcinoma_sscDNA	0.3	112191_SSC-15_Tongue_sscDNA	0.3
94953_HelaS3_Cervical adenocarcinoma_sscDNA	0.2	95017_CAL 27_Squamous cell carcinoma of tongue_sscDNA	0.0

Table IG. Panel 1.3D

	Rel. Exp.(%) Ag2251, Run 159074821	Tissue Name	Rel. Exp.(%) Ag2251, Run 159074821
Liver adenocarcinoma	0.9	Kidney (fetal)	1.9
Pancreas	0.4	Renal ca. 786-0	1.0
Pancreatic ca. CAPAN 2	0.4	Renal ca. A498	4.5
Adrenal gland	0.6	Renal ca. RXF 393	0.0
Thyroid	0.4	Renal ca. ACHN	0.3
Salivary gland	1.2	Renal ca. UO-31	2.8
Pituitary gland	0.7	Renal ca. TK-10	3.8
Brain (fetal)	73.7	Liver	0.0
Brain (whole)	4.6	Liver (fetal)	1.7
Brain (amygdala)	6.4	Liver ca. (hepatoblast) HepG2	1.8
Brain (cerebellum)	1.8	Lung	0.0
Brain (hippocampus)	22.2	Lung (fetal)	3.1
Brain (substantia nigra)	2.1	Lung ca. (small cell) LX-1	4.5

Brain (thalamus)	4.5	Lung ca. (small cell) NCI-H69	8.7
Cerebral Cortex	3.5	Lung ca. (s.cell var.) SHP-77	25.7
Spinal cord	3.2	Lung ca. (large cell) NCI-H460	2.5
glio/astro U87-MG	4.3	Lung ca. (non-sm. cell) A549	2.8
glio/astro U-118-MG	2.2	Lung ca. (non-s.cell) NCI-H23	12.4
astrocytoma SW1783	14.3	Lung ca. (non-s.cell) HOP-62	1.7
neuro*; met SK-N-AS	100.0	Lung ca. (non-s.cl) NCI-H522	28.1
astrocytoma SF-539	0.5	Lung ca. (squam.) SW 900	2.1
astrocytoma SNB-75	13.0	Lung ca. (squam.) NCI-H596	0.7
glioma SNB-19	14.7	Mammary gland	1.0
glioma U251	3.6	Breast ca.* (pl.ef) MCF-7	4.0
glioma SF-295	3.6	Breast ca.* (pl.ef) MDA-MB-231	1.1
Heart (fetal)	3.4	Breast ca.* (pl.ef) T47D	1.1
Heart	0.0	Breast ca. BT-549	16.3
Skeletal muscle (fetal)	15.2	Breast ca. MDA-N	6.4
Skeletal muscle	0.0	Ovary	3.2
Bone marrow	1.8	Ovarian ca. OVCAR-3	1.7
Thymus	21.2	Ovarian ca. OVCAR-4	0.8
Spleen	0.8	Ovarian ca. OVCAR-5	2.3
Lymph node	1.1	Ovarian ca. OVCAR-8	7.3
Colorectal	0.8	Ovarian ca. IGROV-1	2.4
Stomach	0.6	Ovarian ca.* (ascites) SK-OV-3	0.6
Small intestine	2.6	Uterus	0.8
Colon ca. SW480	2.5	Placenta	0.8
Colon ca.* SW620(SW480 met)	1.5	Prostate	1.1
Colon ca. HT29	1.7	Prostate ca.* (bone met) PC-3	3.2
Colon ca. HCT-116	2.4	Testis	69.7
Colon ca. CaCo-2	2.5	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	2.2	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	2.0	Melanoma UACC-62	0.4
Gastric ca.* (liver met) NCI-N87	0.8	Melanoma M14	2.6
Bladder	1.0	Melanoma LOX IMVI	0.7
Trachea	1.8	Melanoma* (met) SK-MEL-5	5.6
Kidney	0.7	Adipose	0.0

Table III. Panel 2D

	Rel. Exp.(%) Ag2251, Run 159075939	Tissue Name	Rel. Exp.(%) Ag2251, Run 159075939
Normal Colon	5.5	Kidney Margin 8120608	0.0
CC Well to Mod Diff (ODO3866)	4.5	Kidney Cancer 8120613	0.0

CC Margin (ODO3866)	2.6	Kidney Margin 8120614	0.6
CC Gr.2 rectosigmoid (ODO3868)	1.2	Kidney Cancer 9010320	0.0
CC Margin (ODO3868)	1.1	Kidney Margin 9010321	1.3
CC Mod Diff (ODO3920)	5.8	Normal Uterus	1.1
CC Margin (ODO3920)	2.3	Uterus Cancer 064011	3.0
CC Gr.2 ascend colon (ODO3921)	4.1	Normal Thyroid	0.6
CC Margin (ODO3921)	0.0	Thyroid Cancer 064010	0.6
CC from Partial Hepatectomy (ODO4309) Mets	1.3	Thyroid Cancer A302152	0.4
Liver Margin (ODO4309)	0.0	Thyroid Margin A302153	2.3
Colon mets to lung (OD04451-01)	4.3	Normal Breast	4.4
Lung Margin (OD04451-02)	0.0	Breast Cancer (OD04566)	1.2
Normal Prostate 6546-1	0.0	Breast Cancer (OD04590-01)	100.0
Prostate Cancer (OD04410)	3.4	Breast Cancer Mets (OD04590-03)	1.5
Prostate Margin (OD04410)	0.0	Breast Cancer Metastasis (OD04655-05)	3.7
Prostate Cancer (OD04720-01)	0.6	Breast Cancer 064006	6.8
Prostate Margin (OD04720-02)	1.8	Breast Cancer 1024	10.4
Normal Lung 061010	5.1	Breast Cancer 9100266	6.6
Lung Met to Muscle (ODO4286)	0.0	Breast Margin 9100265	3.4
Muscle Margin (ODO4286)	0.6	Breast Cancer A209073	7.9
Lung Malignant Cancer (OD03126)	3.9	Breast Margin A209073	2.5
Lung Margin (OD03126)	0.0	Normal Liver	0.0
Lung Cancer (OD04404)	0.0	Liver Cancer 064003	0.6
Lung Margin (OD04404)	0.6	Liver Cancer 1025	0.0
Lung Cancer (OD04565)	0.6	Liver Cancer 1026	0.6
Lung Margin (OD04565)	0.0	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237-01)	99.3	Liver Tissue 6004-N	0.6
Lung Margin (OD04237-02)	2.4	Liver Cancer 6005-T	1.1
Ocular Mel Met to Liver (ODO4310)	0.7	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.0	Normal Bladder	1.8
Melanoma Mets to Lung (OD04321)	18.0	Bladder Cancer 1023	2.8
Lung Margin (OD04321)	0.6	Bladder Cancer A302173	13.2
Normal Kidney	1.4	Bladder Cancer (OD04718-01)	0.0
Kidney Ca, Nuclear grade 2 (OD04338)	8.0	Bladder Normal Adjacent (OD04718-03)	1.3
Kidney Margin (OD04338)	0.0	Normal Ovary	2.8
Kidney Ca Nuclear grade 1/2 (OD04339)	2.4	Ovarian Cancer 064008	4.3
Kidney Margin (OD04339)	0.0	Ovarian Cancer (OD04768-07)	4.0
Kidney Ca, Clear cell type (OD04340)	1.2	Ovary Margin (OD04768-08)	0.0
Kidney Margin (OD04340)	1.0	Normal Stomach	0.8
Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 9060358	0.3

Kidney Margin (OD04348)	0.8	Stomach Margin 9060359	1.2
Kidney Cancer (OD04622-01)	1.1	Gastric Cancer 9060395	0.0
Kidney Margin (OD04622-03)	0.0	Stomach Margin 9060394	1.5
Kidney Cancer (OD04450-01)	4.6	Gastric Cancer 9060397	6.8
Kidney Margin (OD04450-03)	0.6	Stomach Margin 9060396	0.0
Kidney Cancer 8120607	0.6	Gastric Cancer 064005	2.5

**Table II. Panel 4.1D**

	Rel. Exp.(%) Ag2251, Run 244570228	Tissue Name	Rel. Exp.(%) Ag2251, Run 244570228
Secondary Th1 act	12.4	HUVEC IL-1beta	26.4
Secondary Th2 act	16.0	HUVEC IFN gamma	16.5
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	3.3
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	8.9
Secondary Th2 rest	0.0	HUVEC IL-11	39.8
Secondary Tr1 rest	0.0	Lung Microvascular EC none	22.2
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	2.9
Primary Th2 act	11.2	Microvascular Dermal EC none	0.0
Primary Tr1 act	6.7	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	3.3
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	1.6
CD45RA CD4 lymphocyte act	15.3	Coronary artery SMC rest	8.1
CD45RO CD4 lymphocyte act	27.7	Coronary artery SMC TNFalpha + IL-1beta	15.1
CD8 lymphocyte act	0.0	Astrocytes rest	31.6
Secondary CD8 lymphocyte rest	19.9	Astrocytes TNFalpha + IL-1beta	6.4
Secondary CD8 lymphocyte act	4.1	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	2.9
2ry Th1/Th2/Tr1_anti-CD95 CH11	5.2	CCD1106 (Keratinocytes) none	21.6
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	16.3
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	3.2
LAK cells IL-2+ IL-18	2.7	NCI-H292 IL-9	5.8

LAK cells PMA/ionomycin	6.5	NCI-H292 IL-13	5.9
NK Cells IL-2 rest	5.6	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	3.5	HPAEC none	3.3
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	8.2
Two Way MLR 7 day	2.5	Lung fibroblast none	21.3
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	7.7
PBMC PWM	5.3	Lung fibroblast IL-4	10.7
PBMC PHA-L	9.0	Lung fibroblast IL-9	11.2
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	100.0	Lung fibroblast IFN gamma	3.8
B lymphocytes PWM	7.3	Dermal fibroblast CCD1070 rest	14.3
B lymphocytes CD40L and IL-4	14.0	Dermal fibroblast CCD1070 TNF alpha	3.2
EOL-1 dbcAMP	24.8	Dermal fibroblast CCD1070 IL-1 beta	3.7
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	3.7
Dendritic cells none	3.8	Dermal fibroblast IL-4	5.9
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	3.3
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	2.7
Monocytes rest	0.0	Neutrophils rest	6.6
Monocytes LPS	8.2	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	37.4
HUVEC none	30.8	Kidney	0.0
HUVEC starved	59.9		

**Table II. Panel 4D**

Tissue Name	Rel. Exp.(%) g1309, Run 138960659	Rel. Exp.(%) Ag2251, Run 159076647	Tissue Name	Rel. Exp.(%) Ag1309, Run 138960659	Rel. Exp.(%) Ag2251, Run 159076647
Secondary Th1 act	1.5	1.6	HUVEC IL-1beta	1.3	1.6
Secondary Th2 act	1.0	1.2	HUVEC IFN gamma	2.9	2.5
Secondary Tr1 act	2.0	1.7	HUVEC TNF alpha + IFN gamma	1.5	0.1
Secondary Th1 rest	1.7	0.5	HUVEC TNF alpha + IL4	3.5	2.6
Secondary Th2 rest	1.4	0.6	HUVEC IL-11	4.4	1.4
Secondary Tr1 rest	1.4	1.2	Lung Microvascular EC none	1.3	2.3

Primary Th1 act	1.7	2.7	Lung Microvascular EC TNFalpha + IL-1beta	2.1	1.7
Primary Th2 act	3.4	1.9	Microvascular Dermal EC none	6.1	2.6
Primary Tr1 act	5.9	1.2	Microvascular Dermal EC TNFalpha + IL-1beta	2.0	1.3
Primary Th1 rest	12.5	17.1	Bronchial epithelium TNFalpha + IL1beta	2.9	1.6
Primary Th2 rest	6.5	8.6	Small airway epithelium none	0.8	0.4
Primary Tr1 rest	3.7	2.4	Small airway epithelium TNFalpha + IL-1beta	3.1	1.5
CD45RA CD4 lymphocyte act	3.2	1.2	Coronary artery SMC rest	1.3	1.1
CD45RO CD4 lymphocyte act	4.3	2.7	Coronary artery SMC TNFalpha + IL-1beta	1.4	2.0
CD8 lymphocyte act	1.1	1.9	Astrocytes rest	17.8	22.5
Secondary CD8 lymphocyte rest	1.7	0.9	Astrocytes TNFalpha + IL-1beta	6.2	4.7
Secondary CD8 lymphocyte act	1.1	1.0	KU-812 (Basophil) rest	0.3	0.2
CD4 lymphocyte none	1.4	0.5	KU-812 (Basophil) PMA/ionomycin	1.2	0.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	4.5	1.7	CCD1106 (Keratinocytes) none	3.9	3.9
LAK cells rest	1.2	1.6	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	19.5	3.1
LAK cells IL-2	3.1	1.8	Liver cirrhosis	2.0	2.6
LAK cells IL-2+IL-12	1.8	0.7	Lupus kidney	0.3	0.0
LAK cells IL-2+IFN gamma	1.7	1.3	NCI-H292 none	0.7	0.4
LAK cells IL-2+ IL-18	1.5	1.4	NCI-H292 IL-4	1.7	0.4
LAK cells PMA/ionomycin	0.8	0.0	NCI-H292 IL-9	0.0	1.6
NK Cells IL-2 rest	0.9	1.1	NCI-H292 IL-13	0.6	1.6
Two Way MLR 3 day	1.6	2.3	NCI-H292 IFN gamma	0.0	0.3

Two Way MLR 5 day	1.7	0.3	HPAEC none	3.3	2.0
Two Way MLR 7 day	0.8	0.4	HPAEC TNF alpha + IL-1 beta	1.6	0.6
PBMC rest	0.4	0.0	Lung fibroblast none	3.7	3.4
PBMC PWM	6.1	2.1	Lung fibroblast TNF alpha + IL-1 beta	1.6	1.5
PBMC PHA-L	9.9	5.7	Lung fibroblast IL-4	3.6	2.8
Ramos (B cell) none	13.6	6.3	Lung fibroblast IL-9	2.6	3.2
Ramos (B cell) ionomycin	34.9	24.3	Lung fibroblast IL-13	2.7	2.8
B lymphocytes PWM	7.4	7.3	Lung fibroblast IFN gamma	0.5	1.9
B lymphocytes CD40L and IL-4	2.7	4.4	Dermal fibroblast CCD1070 rest	4.2	3.7
EOL-1 dbcAMP	2.6	2.3	Dermal fibroblast CCD1070 TNF alpha	2.4	4.2
EOL-1 dbcAMP PMA/ionomycin	0.3	1.3	Dermal fibroblast CCD1070 IL-1 beta	1.3	2.5
Dendritic cells none	0.5	0.8	Dermal fibroblast IFN gamma	0.7	0.2
Dendritic cells LPS	0.0	0.0	Dermal fibroblast IL-4	0.7	0.8
Dendritic cells anti-CD40	0.3	0.0	IBD Colitis 2	0.2	0.0
Monocytes rest	0.3	0.0	IBD Crohn's	0.0	0.0
Monocytes LPS	1.1	0.3	Colon	3.1	4.2
Macrophages rest	0.6	1.3	Lung	1.5	1.3
Macrophages LPS	0.0	0.0	Thymus	1.6	0.3
HUVEC none	5.3	3.5	Kidney	100.0	100.0
HUVEC starved	12.7	12.4			

**Table IK. general oncology screening panel v 2.4**

Tissue Name	Rel. Exp.(%) Ag2251, Run 259733199	Tissue Name	Rel. Exp.(%) Ag2251, Run 259733199
Colon cancer 1	5.4	Bladder NAT 2	0.0
Colon NAT 1	0.0	Bladder NAT 3	0.2
Colon cancer 2	3.1	Bladder NAT 4	0.0
Colon NAT 2	0.0	Prostate adenocarcinoma 1	4.8
Colon cancer 3	8.8	Prostate adenocarcinoma 2	0.0

Colon NAT 3	1.6	Prostate adenocarcinoma 3	1.1
Colon malignant cancer 4	2.4	Prostate adenocarcinoma 4	7.0
Colon NAT 4	0.0	Prostate NAT 5	1.4
Lung cancer 1	20.4	Prostate adenocarcinoma 6	0.5
Lung NAT 1	0.0	Prostate adenocarcinoma 7	0.6
Lung cancer 2	100.0	Prostate adenocarcinoma 8	0.0
Lung NAT 2	0.6	Prostate adenocarcinoma 9	1.0
Squamous cell carcinoma 3	3.6	Prostate NAT 10	0.4
Lung NAT 3	0.0	Kidney cancer 1	4.8
Metastatic melanoma 1	5.0	Kidney NAT 1	0.3
Melanoma 2	1.8	Kidney cancer 2	5.5
Melanoma 3	2.7	Kidney NAT 2	0.0
Metastatic melanoma 4	22.2	Kidney cancer 3	4.9
Metastatic melanoma 5	17.8	Kidney NAT 3	1.2
Bladder cancer 1	0.7	Kidney cancer 4	0.0
Bladder NAT 1	0.0	Kidney NAT 4	0.0
Bladder cancer 2	0.0		

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag2251 Highest expression of this gene is detected in orthoarthritis bone (CT=31.6). Low expression of this gene is also seen in in samples derived from normal and orthoarthritis bone, synovium samples, from normal lung, COPD lung, emphysema, atopic asthma, asthma, allergy, Crohn's disease (normal matched control and diseased), ulcerative colitis(normal matched control and diseased), and psoriasis (normal matched control and diseased). Therefore, therapeutic modulation of this gene product may ameliorate symptoms/conditions associated with autoimmune and inflammatory disorders including psoriasis, allergy, asthma, inflammatory bowel diseases (Crohn's and ulcerative colitis), and osteoarthritis.

**CNS\_neurodegeneration\_v1.0 Summary:** Ag2251 Highest expression of this gene in this panel is detected in the cerebral cortex of an Alzheimer's patient (CT=32.7). While no association between the expression of this gene and the presence of Alzheimer's disease is detected in this panel, these results confirm the expression of this gene in areas that degenerate in Alzheimer's disease. Please see Panel 1.3D and 1.5 for a discussion of potential utility of this gene in the central nervous system.

**General\_screening\_panel\_v1.5 Summary:** Ag2251 Highest expression of this gene is detected in fetal brain (CT=27.1). Low expression of this gene is also seen all the regions of adult brain including amygdala, hippocampus, substantia nigra, thalamus,

cerebellum, cerebral cortex, and spinal cord. Interestingly, expression of this gene is higher in fetal compared to the adult whole brain (32.6). This gene represents the human ortholog of cerebroglycan, a glycosylphosphatidylinositol (GPI)-anchored HSPG that is found in the developing rat brain. Heparan sulfate proteoglycans (HSPGs) are found on the surface of  
5 all adherent cells and participate in the binding of growth factors, extracellular matrix glycoproteins, cell adhesion molecules, and proteases and antiproteases. Unlike other known integral membrane HSPGs, including glypican and members of the syndecan family of transmembrane proteoglycans, cerebroglycan is apparently expressed in only one tissue in the rat: the nervous system and it is really present only during fetal development in  
10 immature neurons. Expression of this gene in human fetal and all the regions of adult brain regions suggest that this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

In addition, significant expression of this gene is also seen in number of cancer cell  
15 lines derived from pancreatic, gastric, colon, lung, renal, breast, ovarian, prostate, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Expression of this gene is higher in cell lines derived especially from breast and lung cancers and also in fetal tissues including lung, heart, kidney and skeletal muscle (CTs=30-32.7) compared to respective adult tissues  
20 (CTs=33-35.5). Thus, this gene may play role growth or development of the cells, especially during tumorigenesis and may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the expression or function of this gene through the use of antibodies may be effective in the treatment of these cancers, especially breast and lung cancers.

25 **Oncology\_cell\_line\_screening\_panel\_v3.2 Summary:** Ag2251 Highest expression of this gene is detected in small cell lung cancer DMS-79 cell line (CT=28.7). High expression of this gene is seen in number of cell lines derived from lung cancer. Moderate to low expression of this gene is also seen in number of cell lines derived from brain, colon, cervical, bladder and bone cancers, T and B cell lymphomas. Please see panel  
30 1.5 and 1.3D for further discussion on the utility of this gene.

**Panel 1.3D Summary:** Ag2251 The highest level of expression of this gene is seen in a CNS cancer cell line SK-N-AS (CT=29.6). The gene is also expressed at higher levels

in cell lines derived from lung, prostate, and breast cancers compared to the normal tissues and may play a role in these cancers. Thus, expression of this gene could be used as a marker or as a therapeutic for lung, prostate and breast cancer. In addition, therapeutic modulation of the activity of the product of this gene, through the use of peptides,  
5 antibodies, chimeric molecules or small molecule drugs, may be useful in the treatment of these cancers.

This gene is also expressed at higher levels in fetal liver, lung, skeletal muscle, and heart (CTs=32-35) when compared to the expression in adult tissues (CTs=40). These results suggest that expression of this gene could potentially be used to distinguish between  
10 the adult and fetal phenotypes of these tissues. Furthermore, the difference in expression in fetal and adult tissue may also indicate an involvement of the gene product in the differentiation processes leading to the formation of the adult organs. Therefore, the protein encoded by this gene could potentially play a role in the regeneration of these tissues in the adult.

15 This gene, a glypican homolog, is expressed at moderate to low levels across many regions of the brain. These regions include the hippocampus, amygdala, thalamus and cerebral cortex, all of which are key regions subject to Alzheimer's disease neurodegeneration. Furthermore, glypican is expressed in senile plaques and neurofibrillary tangles, also indicating a role in Alzheimer's disease. Therefore, the expression profile of  
20 this gene suggests that antibodies against the protein encoded by this gene can be used to distinguish neurodegenerative disease in the human brain. Furthermore, since glypican are components of senile plaques which are thought to give rise to the dementia pathology of Alzheimer's disease, agents that target this gene and disrupt its role in senile plaques (Ref. 1) may have utility in treating the cause and symptoms of Alzheimer's disease as well  
25 as other neurodegenerative diseases that involve this glypican.

#### Reference:

1. Verbeek MM, Otte-Holler I, van den Born J, van den Heuvel LP, David G, Wesseling P, de Waal RM. (1999) Agrin is a major heparan sulfate proteoglycan accumulating in Alzheimer's disease brain. *Am J Pathol.* 155:2115-25. PMID: 10595940

**Panel 2D Summary:** Ag2251 The highest expression of this gene is seen in a breast cancer sample (CT = 30.3). The expression of this gene appears to show an association with samples derived from colon, lung, kidney, breast, bladder and gastric cancers when compared to the matched normal tissue. Thus, expression of this gene could be used as a marker for these cancers. In addition, therapeutic activity of the product of this gene, through the use of peptides, antibodies, chimeric molecules or small molecule drugs, may be useful in the treatment of colon, lung, kidney, breast, bladder and gastric cancers.

**Panel 4.1D Summary:** Ag2251 Highest expression of this gene is seen in ionomycin activated Ramos B cells (CT=31.4). Expression of this gene is low or undetectable in resting Ramos B cells (CT=40). B cells represent a principle component of immunity and contribute to the immune response in a number of important functional roles, including antibody production. Production of antibodies against self-antigens is a major component in autoimmune disorders. In addition, low expression of this gene is also seen in eosinophils, HUVEC cells, activated secondary Th1 and Th2 cells, naive and memory T cells, lung and dermal fibroblast and thymus. Therefore, therapeutic modulation of this gene product may reduce or eliminate the symptoms of patients suffering from asthma, allergies, chronic obstructive pulmonary disease, emphysema, Crohn's disease, ulcerative colitis, rheumatoid arthritis, psoriasis, osteoarthritis, systemic lupus erythematosus and other autoimmune disorders.

**Panel 4D Summary:** Ag2251/Ag1309 Two experiments using two different probe and primer sets produce results that are in very good agreement, with highest expression seen in the kidney (CTs=28-29). This high level of expression in the kidney suggests that expression of this gene can serve as a marker for kidney tissue. This gene is also expressed at low level in activated Ramos B cell line, in activated primary B cells, Th1 T cells, activated HUVEC and keratinocytes. This gene encodes for a protein that is a homolog of a glypican molecule, which belongs to the family of HSPG (heparan sulfate proteoglycans). Glypicans can regulate the activity of a wide variety of growth and survival factors. Therefore, therapeutic modulation of the expression or function of this gene or gene product, through the use of antibody drugs could potentially prevent T and B cell activation in the treatment of autoimmune mediated diseases such as insulin-dependent diabetes mellitus, rheumatoid arthritis, Crohn's disease, allergies, delayed type hypersensitivity, asthma, and psoriasis.

general oncology screening panel\_v\_2.4 Summary: Ag2251 Highest expression of this gene is detected in lung cancer2 (CT=30.7). Moderate to low expression of this gene is also seen in lung cancer1, two metastatic melanoma and prostate cancer samples. Therefore, expression of this gene may be used as a diagnostic marker to detect the presence of these cancers and also, therapeutic modulation of this gene or its product through the use of antibodies or small molecule drug may be useful in the treatment of metastatic melanoma, lung and prostate cancers.

**J. CG54443-07 (NOV16b): CG8841 PROTEIN-like protein.**

Expression of gene CG54443-07 was assessed using the primer-probe sets Ag2000 and Ag6688, described in Tables JA and JB. Results of the RTQ-PCR runs are shown in Tables JC, JD and JE.

**Table JA. Probe Name Ag2000**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-actccaccaagaagatccagtt-3'	22	527	153
Probe	TET-5'-tctcttctggaagctctgcgacttca-3'-TAMRA	26	567	154
Reverse	5'-gcacgaagaagaggaatttctt-3'	22	595	155

**Table JB. Probe Name Ag6688**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ccaccaagacgcagc-3'	15	185	156
Probe	TET-5'-aagccaccgatgatgcctatg-3'-TAMRA	21	206	157
Reverse	5'-gagcaggtggtttaggg-3'	18	247	158

**Table JC. Panel 1.3D**

Tissue Name	Rel. Exp.(%) g2000, Run 147805564	Tissue Name	Rel. Exp.(%) Ag2000, Run 147805564
Liver adenocarcinoma	9.8	Kidney (fetal)	6.0
Pancreas	24.8	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	1.3	Renal ca. A498	1.0
Adrenal gland	3.3	Renal ca. RXF 393	0.0

Thyroid	11.0	Renal ca. ACHN	1.5
Salivary gland	30.6	Renal ca. UO-31	1.1
Pituitary gland	30.4	Renal ca. TK-10	2.4
Brain (fetal)	13.0	Liver	0.7
Brain (whole)	39.2	Liver (fetal)	2.5
Brain (amygdala)	23.7	Liver ca. (hepatoblast) HepG2	8.8
Brain (cerebellum)	21.0	Lung	12.9
Brain (hippocampus)	46.7	Lung (fetal)	30.4
Brain (substantia nigra)	10.4	Lung ca. (small cell) LX-1	8.7
Brain (thalamus)	33.2	Lung ca. (small cell) NCI-H69	29.5
Cerebral Cortex	100.0	Lung ca. (s.cell var.) SHP-77	33.0
Spinal cord	14.6	Lung ca. (large cell) NCI-H460	0.9
glio/astro U87-MG	0.1	Lung ca. (non-sm. cell) A549	15.9
glio/astro U-118-MG	0.3	Lung ca. (non-s.cell) NCI-H23	2.3
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	3.3
neuro*; met SK-N-AS	4.3	Lung ca. (non-s.cl) NCI-H522	1.8
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	20.2
astrocytoma SNB-75	35.6	Lung ca. (squam.) NCI-H596	3.3
glioma SNB-19	5.7	Mammary gland	40.1
glioma U251	2.1	Breast ca.* (pl.ef) MCF-7	42.0
glioma SF-295	2.6	Breast ca.* (pl.ef) MDA-MB-231	6.3
Heart (fetal)	44.4	Breast ca.* (pl.ef) T47D	73.2
Heart	3.6	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	69.3	Breast ca. MDA-N	0.2
Skeletal muscle	0.6	Ovary	17.6
Bone marrow	1.8	Ovarian ca. OVCAR-3	23.5
Thymus	2.9	Ovarian ca. OVCAR-4	9.2
Spleen	14.8	Ovarian ca. OVCAR-5	13.0
Lymph node	8.6	Ovarian ca. OVCAR-8	2.8
Colorectal	18.9	Ovarian ca. IGROV-1	1.9
Stomach	68.3	Ovarian ca.* (ascites) SK-OV-3	2.7
Small intestine	21.9	Uterus	9.9
Colon ca. SW480	10.0	Placenta	27.2
Colon ca.* SW620(SW480 met)	2.9	Prostate	25.9
Colon ca. HT29	16.8	Prostate ca.* (bone met) PC-3	18.7
Colon ca. HCT-116	5.5	Testis	7.4
Colon ca. CaCo-2	11.6	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	27.0	Melanoma* (met) Hs688(B).T	0.1
Colon ca. HCC-2998	17.2	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	48.6	Melanoma M14	0.0
Bladder	10.7	Melanoma LOX IMVI	0.0
Trachea	36.1	Melanoma* (met) SK-MEL-5	0.7
Kidney	1.9	Adipose	4.3

**Table JD. Panel 2.2**

	Rel. Exp.(%) Ag2000, Run 174232799	Tissue Name	Rel. Exp.(%) Ag2000, Run 174232799
Normal Colon	14.0	Kidney Margin (OD04348)	10.7
Colon cancer (OD06064)	21.3	Kidney malignant cancer (OD06204B)	29.7
Colon Margin (OD06064)	24.5	Kidney normal adjacent tissue (OD06204E)	3.8
Colon cancer (OD06159)	7.0	Kidney Cancer (OD04450-01)	4.1
Colon Margin (OD06159)	11.0	Kidney Margin (OD04450-03)	5.0
Colon cancer (OD06297-04)	8.7	Kidney Cancer 8120613	1.3
Colon Margin (OD06297-05)	14.1	Kidney Margin 8120614	7.6
CC Gr.2 ascend colon (ODO3921)	9.4	Kidney Cancer 9010320	2.7
CC Margin (ODO3921)	4.8	Kidney Margin 9010321	2.9
Colon cancer metastasis (OD06104)	3.1	Kidney Cancer 8120607	9.0
Lung Margin (OD06104)	10.2	Kidney Margin 8120608	3.0
Colon mets to lung (OD04451-01)	10.8	Normal Uterus	9.0
Lung Margin (OD04451-02)	8.3	Uterine Cancer 064011	4.9
Normal Prostate	42.9	Normal Thyroid	5.4
Prostate Cancer (OD04410)	17.2	Thyroid Cancer 064010	2.8
Prostate Margin (OD04410)	10.4	Thyroid Cancer A302152	6.3
Normal Ovary	7.6	Thyroid Margin A302153	4.6
Ovarian cancer (OD06283-03)	9.5	Normal Breast	19.6
Ovarian Margin (OD06283-07)	4.7	Breast Cancer (OD04566)	15.8
Ovarian Cancer 064008	7.3	Breast Cancer 1024	22.4
Ovarian cancer (OD06145)	0.4	Breast Cancer (OD04590-01)	47.6
Ovarian Margin (OD06145)	7.3	Breast Cancer Mets (OD04590-03)	41.2
Ovarian cancer (OD06455-03)	18.0	Breast Cancer Metastasis (OD04655-05)	100.0
Ovarian Margin (OD06455-07)	2.4	Breast Cancer 064006	11.1
Normal Lung	18.6	Breast Cancer 9100266	49.0
Invasive poor diff. lung adeno (ODO4945-01)	10.0	Breast Margin 9100265	20.7
Lung Margin (ODO4945-03)	5.7	Breast Cancer A209073	18.6
Lung Malignant Cancer (OD03126)	17.6	Breast Margin A2090734	21.5
Lung Margin (OD03126)	3.9	Breast cancer (OD06083)	81.2
Lung Cancer (OD05014A)	11.3	Breast cancer node metastasis (OD06083)	66.0
Lung Margin (OD05014B)	0.2	Normal Liver	2.4
Lung cancer (OD06081)	4.2	Liver Cancer 1026	4.4
Lung Margin (OD06081)	6.3	Liver Cancer 1025	4.6
Lung Cancer (OD04237-01)	4.6	Liver Cancer 6004-T	3.8

Lung Margin (OD04237-02)	9.1	Liver Tissue 6004-N	1.5
Ocular Melanoma Metastasis	0.7	Liver Cancer 6005-T	12.1
Ocular Melanoma Margin (Liver)	2.8	Liver Tissue 6005-N	9.6
Melanoma Metastasis	0.3	Liver Cancer 064003	1.5
Melanoma Margin (Lung)	9.2	Normal Bladder	19.6
Normal Kidney	2.5	Bladder Cancer 1023	6.3
Kidney Ca, Nuclear grade 2 (OD04338)	9.7	Bladder Cancer A302173	8.7
Kidney Margin (OD04338)	1.7	Normal Stomach	62.4
Kidney Ca Nuclear grade 1/2 (OD04339)	4.2	Gastric Cancer 9060397	5.1
Kidney Margin (OD04339)	4.1	Stomach Margin 9060396	38.4
Kidney Ca, Clear cell type (OD04340)	2.7	Gastric Cancer 9060395	21.5
Kidney Margin (OD04340)	6.7	Stomach Margin 9060394	43.5
Kidney Ca, Nuclear grade 3 (OD04348)	0.6	Gastric Cancer 064005	11.4

**Table JE. Panel 4D**

	Rel. Exp.(%) Ag2000, Run 165822435	Tissue Name	Rel. Exp.(%) Ag2000, Run 165822435
Secondary Th1 act	0.2	HUVEC IL-1beta	4.8
Secondary Th2 act	0.3	HUVEC IFN gamma	14.7
Secondary Tr1 act	0.6	HUVEC TNF alpha + IFN gamma	1.9
Secondary Th1 rest	0.1	HUVEC TNF alpha + IL4	4.0
Secondary Th2 rest	0.7	HUVEC IL-11	15.6
Secondary Tr1 rest	0.3	Lung Microvascular EC none	14.4
Primary Th1 act	0.1	Lung Microvascular EC TNFalpha + IL-1beta	6.3
Primary Th2 act	0.2	Microvascular Dermal EC none	15.4
Primary Tr1 act	0.1	Microvascular Dermal EC TNFalpha + IL-1beta	5.1
Primary Th1 rest	0.4	Bronchial epithelium TNFalpha + IL1beta	2.6
Primary Th2 rest	0.2	Small airway epithelium none	0.8
Primary Tr1 rest	0.1	Small airway epithelium TNFalpha + IL-1beta	3.4
CD45RA CD4 lymphocyte act	0.3	Coronary artery SMC rest	0.1
CD45RO CD4 lymphocyte act	0.7	Coronary artery SMC TNFalpha + IL-1beta	0.3
CD8 lymphocyte act	0.8	Astrocytes rest	3.6
Secondary CD8 lymphocyte rest	0.9	Astrocytes TNFalpha + IL-1beta	6.9

Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	2.7	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.3
LAK cells rest	1.4	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.5
LAK cells IL-2	2.6	Liver cirrhosis	11.2
LAK cells IL-2+IL-12	1.4	Lupus kidney	9.2
LAK cells IL-2+IFN gamma	1.2	NCI-H292 none	20.2
LAK cells IL-2+ IL-18	1.6	NCI-H292 IL-4	17.4
LAK cells PMA/ionomycin	0.3	NCI-H292 IL-9	21.6
NK Cells IL-2 rest	0.4	NCI-H292 IL-13	9.5
Two Way MLR 3 day	1.2	NCI-H292 IFN gamma	10.3
Two Way MLR 5 day	0.4	HPAEC none	13.7
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	9.2
PBMC rest	0.8	Lung fibroblast none	0.2
PBMC PWM	0.2	Lung fibroblast TNF alpha + IL-1 beta	0.8
PBMC PHA-L	0.3	Lung fibroblast IL-4	0.1
Ramos (B cell) none	0.5	Lung fibroblast IL-9	0.2
Ramos (B cell) ionomycin	0.7	Lung fibroblast IL-13	0.2
B lymphocytes PWM	0.8	Lung fibroblast IFN gamma	0.3
B lymphocytes CD40L and IL-4	5.8	Dermal fibroblast CCD1070 rest	0.1
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.1
Dendritic cells none	0.2	Dermal fibroblast IFN gamma	0.1
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.1
Dendritic cells anti-CD40	0.0	IBD Colitis 2	2.9
Monocytes rest	0.0	IBD Crohn's	9.2
Monocytes LPS	0.0	Colon	100.0
Macrophages rest	0.1	Lung	19.3
Macrophages LPS	0.0	Thymus	11.1
HUVEC none	7.3	Kidney	6.4
HUVEC starved	17.7		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag6688 Expression of this gene is limited to a single sample from the parietal cortex (CT=34).

**General\_screening\_panel\_v1.6 Summary:** Ag6688 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

**Panel 1.3D Summary:** Ag2000 Highest expression of this gene, a homolog of a transmembrane multi-pass protein, is seen in the cerebral cortex (CT=26.8), with moderate

expression detectable across all regions of the brain. Because this gene shows a large down-regulation in brain cancers, its absence would be an excellent marker to determine if brain tissue was pre-cancerous in the examining and classifying of postmortem tissue

Expression of this gene is also widespread among tissues with metabolic relevance, including adipose, pancreas, adult and fetal heart, adult and fetal liver, adult and fetal skeletal muscle, and the adrenal, pituitary, and thyroid glands. The gene is expressed at much higher levels in fetal heart and skeletal muscle (CTs=28) than in adult heart and skeletal muscle (CTs=31-34). This differential expression pattern suggests that this gene expression could be used to differentiate between the two tissue sources for heart and skeletal muscle. Furthermore, the significantly higher level of expression of the gene in fetal skeletal muscle suggests that this gene product may be involved in muscular growth or development in the fetus and could potentially act in a regenerative capacity in an adult. Therefore, therapeutic modulation of this gene could be useful in the treatment of muscle related diseases and the treatment of weak or dystrophic muscle.

This gene is also expressed at significant levels in cell lines derived from ovarian, breast, lung, gastric, prostate and colon cancers compared to the normal tissues. Thus, the expression of this gene could be of use as a marker or as a therapeutic for ovarian, breast, lung, gastric, prostate and colon. In addition, therapeutic modulation of the product of this gene, through the use of peptides, chimeric molecules or small molecule drugs, may be useful in the treatment of these cancers.

**Panel 2.2 Summary:** Ag2000 Highest expression of this gene is seen in breast cancer (CT=28) as is seen in Panel 1.3D. In addition, there is significant overexpression of this gene in a cluster of breast, lung, and ovarian cancer samples when compared to corresponding normal tissues. Thus, expression of this gene could be used to differentiate breast, ovarian and lung cancers from normal tissue and as a marker for the presence of these cancers. Furthermore, therapeutic modulation of the protein product of this gene could be beneficial in the treatment of breast, ovarian and lung cancers. The expression of this gene also shows a reverse association with some normal stomach samples when compared to the matched gastric cancer tissue. This suggests that the this gene could be used to distinguish between normal and cancerous gastric tissue and that therapeutic modulation of the gene product may be useful in the treatment of gastric cancer.

**Panel 4.1D Summary:** Ag6688 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

**Panel 4D Summary:** Ag2000 The highest expression of this gene is found in the colon (CT=26.2), with modest expression detectable in the muco-epidermoid cell line H292, and the lung. It is also expressed at moderate levels on HUVEC and lung microvasculature regardless of their activation status. The protein encoded by this gene is homologous to an epidermal growth factor related protein (fibropellin like) and could be used as a marker of lung muco-epidermoid cells, colon or vasculature. The putative protein encoded by the transcript may also play an important role in the normal homeostasis of these tissues. Small molecule or antibody therapeutics designed with this gene product could be important for maintaining or restoring normal function to these organs during inflammation associated with asthma and emphysema.

**K. CG58495-03 (NOV 17b): Pulmonary surfactant-associated protein A precursor.**

Expression of gene CG58495-03 was assessed using the primer-probe set Ag7945, described in Table KA.

**Table KA. Probe Name Ag7945**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gcgtgcgaagtgaagga-3'	17	135	159
Probe	TET-5'-ctccaagccacactccacgacttcag-3'-TAMRA	26	153	160
Reverse	5'-ctgagggctccccttgtc-3'	18	194	161

**CNS\_neurodegeneration\_v1.0 Summary:** Ag7945 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

**Panel 4.1D Summary:** Ag7945 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

**L. CG97482-02 (NOV18b): S100 Calcium-Binding Protein-like.**

Expression of gene CG97482-02 was assessed using the primer-probe set Ag6384, described in Table LA. Results of the RTQ-PCR runs are shown in Table LB. Please note that CG97482-02 represents a full length physical clone.

**5 Table LA. Probe Name Ag6384**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tggccctcatcgacgt-3'	16	44	162
Probe	TET-5'-agctcatcaacaatgagctttccatt-3'-TAMRA	27	122	163
Reverse	5'-gcagtagtaaccacaacctcct-3'	22	170	164

**Table LB. CNS neurodegeneration v1.0**

Tissue Name	Rel. Exp.(%) Ag6384, Run 269253944	Tissue Name	Rel. Exp.(%) Ag6384, Run 269253944
AD 1 Hippo	28.3	Control (Path) 3 Temporal Ctx	5.1
AD 2 Hippo	92.7	Control (Path) 4 Temporal Ctx	25.9
AD 3 Hippo	4.2	AD 1 Occipital Ctx	19.5
AD 4 Hippo	17.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	36.1	AD 3 Occipital Ctx	7.7
AD 6 Hippo	100.0	AD 4 Occipital Ctx	37.1
Control 2 Hippo	82.9	AD 5 Occipital Ctx	68.3
Control 4 Hippo	65.1	AD 6 Occipital Ctx	39.2
Control (Path) 3 Hippo	13.7	Control 1 Occipital Ctx	9.7
AD 1 Temporal Ctx	22.4	Control 2 Occipital Ctx	61.6
AD 2 Temporal Ctx	74.7	Control 3 Occipital Ctx	35.8
AD 3 Temporal Ctx	4.7	Control 4 Occipital Ctx	47.3
AD 4 Temporal Ctx	42.0	Control (Path) 1 Occipital Ctx	97.3
AD 5 Inf Temporal Ctx	68.3	Control (Path) 2 Occipital Ctx	25.5
AD 5 Sup Temporal Ctx	75.8	Control (Path) 3 Occipital Ctx	10.4
AD 6 Inf Temporal Ctx	57.0	Control (Path) 4 Occipital Ctx	12.1
AD 6 Sup Temporal Ctx	45.4	Control 1 Parietal Ctx	26.2
Control 1 Temporal Ctx	21.8	Control 2 Parietal Ctx	46.3
Control 2 Temporal Ctx	56.3	Control 3 Parietal Ctx	51.1
Control 3 Temporal Ctx	35.4	Control (Path) 1 Parietal Ctx	58.2
Control 3 Temporal Ctx	30.6	Control (Path) 2 Parietal Ctx	59.9
Control (Path) 1 Temporal Ctx	42.0	Control (Path) 3 Parietal Ctx	6.2

Control (Path) 2 Temporal Ctx	39.8	Control (Path) 4 Parietal Ctx	37.9
-------------------------------	------	-------------------------------	------

**CNS\_neurodegeneration\_v1.0 Summary:** Ag6384 No differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. However, this panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

**Panel 4.1D Summary:** Ag6384 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

#### **Example D: Identification of Single Nucleotide Polymorphisms in NOVX nucleic acid sequences**

Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern. Examples include alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, and stability of transcribed message.

SeqCalling assemblies produced by the exon linking process were selected and extended using the following criteria. Genomic clones having regions with 98% identity to

all or part of the initial or extended sequence were identified by BLASTN searches using the relevant sequence to query human genomic databases. The genomic clones that resulted were selected for further analysis because this identity indicates that these clones contain the genomic locus for these SeqCalling assemblies. These sequences were  
5 analyzed for putative coding regions as well as for similarity to the known DNA and protein sequences. Programs used for these analyses include Grail, Genscan, BLAST, HMMER, FASTA, Hybrid and other relevant programs.

Some additional genomic regions may have also been identified because selected SeqCalling assemblies map to those regions. Such SeqCalling sequences may have  
10 overlapped with regions defined by homology or exon prediction. They may also be included because the location of the fragment was in the vicinity of genomic regions identified by similarity or exon prediction that had been included in the original predicted sequence. The sequence so identified was manually assembled and then may have been extended using one or more additional sequences taken from CuraGen Corporation's human  
15 SeqCalling database. SeqCalling fragments suitable for inclusion were identified by the CuraTools™ program SeqExtend or by identifying SeqCalling fragments mapping to the appropriate regions of the genomic clones analyzed.

The regions defined by the procedures described above were then manually integrated and corrected for apparent inconsistencies that may have arisen, for example,  
20 from miscalled bases in the original fragments or from discrepancies between predicted exon junctions, EST locations and regions of sequence similarity, to derive the final sequence disclosed herein. When necessary, the process to identify and analyze SeqCalling assemblies and genomic clones was reiterated to derive the full length sequence (Alderborn et al., Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate  
25 DNA Sequencing. Genome Research. 10 (8) 1249-1265, 2000).

Variants are reported individually but any combination of all or a select subset of variants are also included as contemplated NOVX embodiments of the invention.

#### **NOV1b SNP Data (CG108030-02).**

30 Seven polymorphic variants of NOV1b have been identified and are shown in Table 19A.

**Table 19A. Variant of NOV1b.**

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381876	354	T	C	116	Leu	Pro
13381877	627	T	C	207	Leu	Pro
13381845	2249	A	G	0	N/A	N/A
13381844	2454	C	T	0	N/A	N/A
13381881	2949	T	C	0	N/A	N/A
13381882	2959	A	G	0	N/A	N/A
13381883	3124	A	G	0	N/A	N/A

5 **NOV2d SNP Data (CG115907-02).**

Four polymorphic variants of NOV2d have been identified and are shown in Table 19B.

**Table 19B. Variant of NOV2d.**

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381868	204	T	C	25	Thr	Thr
13381869	1892	T	A	588	Leu	His
13381842	2131	C	A	668	Pro	Thr
13381871	2544	G	A	805	Leu	Leu

10

**NOV6a SNP Data (CG155653-01).**

Three polymorphic variants of NOV6a have been identified and are shown in Table 15 19C.

**Table 19C. Variant of NOV6a.**

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381864	301	G	A	42	Gly	Ser
13381889	1260	G	T	361	Arg	Ser
13381867	4013	G	A	0	N/A	N/A

5 **NOV7a SNP Data (CG160093-01).**

Three polymorphic variants of NOV7a have been identified and are shown in Table 19D.

**Table 19D. Variant of NOV7a.**

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381888	966	A	G	315	Glu	Gly
13381887	980	A	G	320	Thr	Ala
13381886	1008	T	C	329	Leu	Ser

10

**NOV9a SNP Data (CG165528-01).**

Four polymorphic variants of NOV9a have been identified and are shown in Table 19E.

15 **Table 19E. Variant of NOV9a.**

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381837	78	C	T	0	N/A	N/A
13381838	4640	T	C	1514	Val	Ala
13381839	4754	A	G	0	N/A	N/A
13381840	4936	A	G	0	N/A	N/A

**NOV12d SNP Data (CG165719-01).**

Four polymorphic variants of NOV12d have been identified and are shown in Table 19F.

**Table 19F. Variant of NOV12d.**

5

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381873	291	C	A	77	Ser	Ser
13381875	475	T	C	139	Phe	Leu
13381874	559	G	A	167	Ala	Thr
13381884	631	T	C	191	Phe	Leu

**NOV17b SNP Data (CG58495-03).**

Three polymorphic variants of NOV17b have been identified and are shown in Table 19G.

10

**Table 19G. Variant of NOV17b.**

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13376633	151	A	G	24	Glu	Gly
13381911	386	T	C	102	Tyr	Tyr
13376634	501	A	C	141	Lys	Gln

15

**NOV18b SNP Data (CG97482-02).**

One polymorphic variant of NOV18b has been identified and is shown in Table 19H.

**Table 19H. Variant of NOV18b.**

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13376808	176	T	C	53	Val	Ala

5 **Example E. CG50970-01, NOV15b.**

**Role in inflammation:** This transcript encodes glypican 2 a glycosylphosphatidylinositol (gpi) anchored cell surface heparan sulfate proteoglycan. This type of proteoglycan can bind cytokines and is potentially involved in lymphocytic migration and activation (1). Additionally, this molecule is also found in bone marrow and cartilage (2-3) and may be involved in osteoblast function and hematopoiesis.

Therapeutic function: Antibody therapeutics which antagonized the function of the protein encoded for by this transcript could be used to reduce or inhibit lymphocyte extravasation associated with inflammation due to asthma, emphysema, rheumatoid arthritis, IBD or psoriasis. Antibodies may also block tissue changes associated with osteoarthritis (4).

**Example E1: Gene Expression analysis using CuraChip in human tissues from tumors and from equivalent normal tissues**

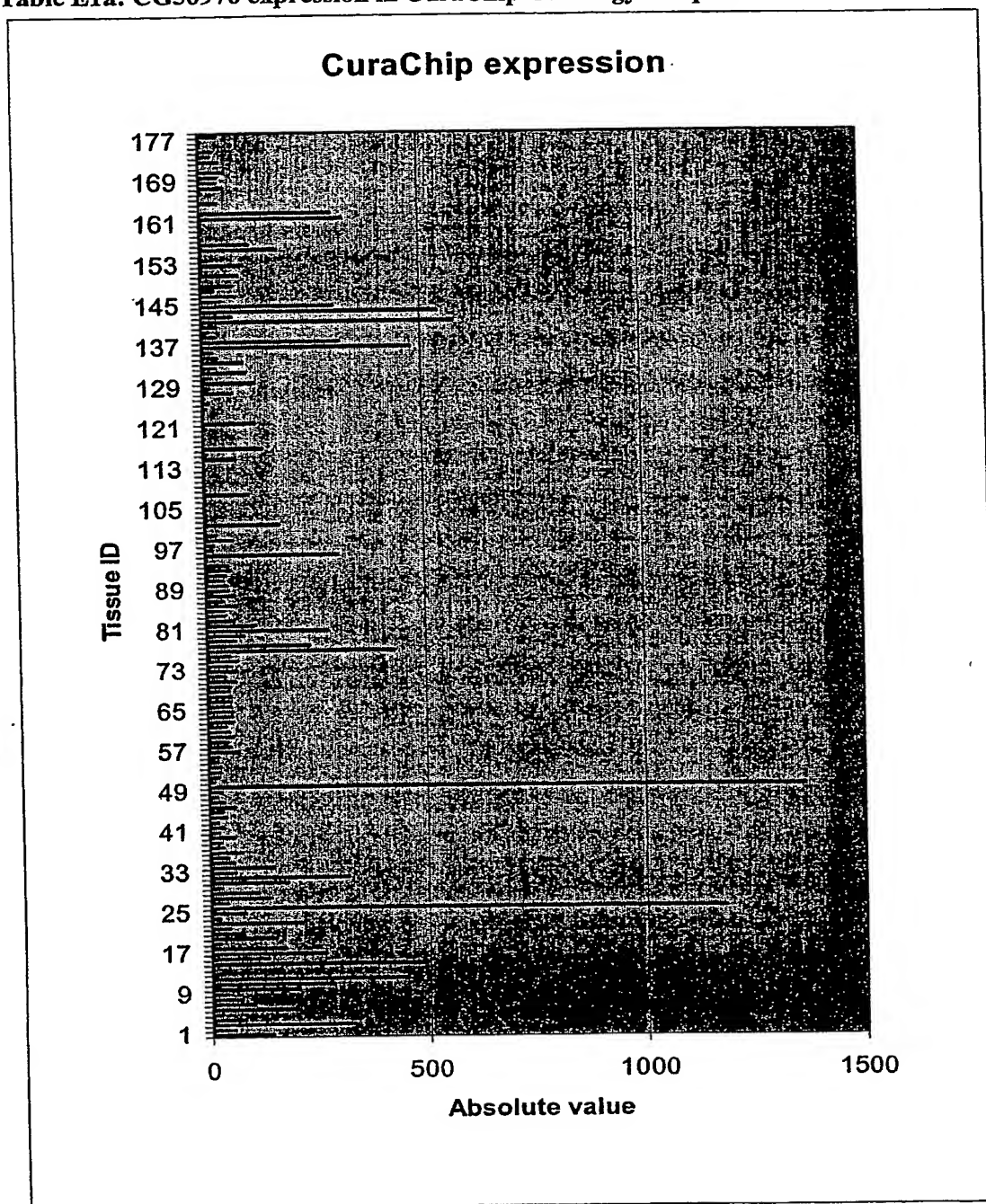
20 CuraGen has developed a gene microarray (CuraChip 1.2) for target identification. It provides a high-throughput means of global mRNA expression analyses of CuraGen's collection of cDNA sequences representing the Pharmaceutically Tractable Genome (PTG). This sequence set includes genes which can be developed into protein therapeutics, or used to develop antibody or small molecule therapeutics. CuraChip 1,2 contains ~11,000 oligos representing approximately 8,500 gene loci, including (but not restricted to) kinases, ion channels, G-protein coupled receptors (GPCRs), nuclear hormone receptors, proteases, transporters, metabolic enzymes, hormones, growth factors, chemokines, cytokines, complement and coagulation factors, and cell surface receptors.

30

The CuraChip cDNAs were represented as 30-mer oligodeoxyribonucleotides (oligos) on a glass microchip. Hybridization methods using the longer CuraChip oligos are more specific compared to methods using 25-mer oligos. CuraChip oligos were synthesized with a linker, purified to remove truncated oligos (which can influence hybridization strength and specificity), and spotted on a glass slide. Oligo-dT primers were used to generate cRNA probes for hybridization from samples of interest. A biotin-avidin conjugation system was used to detect hybridized probes with a fluorophore-labeled secondary antibody. Gene expression was analyzed using clustering and correlation bioinformatics tools such as Spotfire® (Spotfire, Inc., 212 Elm Street, Somerville, MA 02144) and statistical tools such as multivariate analysis (MVA).

**Results of PTG Chip 1.2:** One hundred seventy-eight samples of RNA from tissues obtained from surgically dissected tumors, non-diseased tissues from the corresponding organs and tumor xenografts grown in nude nu/nu mice were used to generate probes and run on PTG Chip 1.2. An oligo (optg2\_0011299) that corresponds to CG50970 on the PTG Chip 1.2 was scrutinized for its expression profile. The statistical analysis identify significant over-expression in a subset of lung tumors, about 30%, compared with corresponding normal lung tissue and strong expression in breast cancers, also about 30%, which do not have matched normal tissue. It is also useful that the expression of this gene is maintained when human tumor cell lines are grown as tumor xenografts in nude mice, especially by the lung tumor cell lines NCI-H82 and NCI-H69. Therefore these tumor xenografts can be used as animal models.

Thus, based upon its profile, the expression of this gene could be of use as a marker for subsets of lung and breast cancers. In addition, therapeutic inhibition of the activity of the product of this gene, through the use of antibodies or small molecule drugs, may be useful in the therapy of lung and breast cancers that express CG50970.

**Table E1a: CG50970 expression in CuraChip Oncology samples**

**Example E2. Protein expression and purification**

CG50970-05 is expressed and purified in the CHO stable cell system using the Wave bioreactor.

To separate the glycanated form of the proteoglycan from the unglycanated core protein, the conditioned medium was applied to a  $0.9 \times 8$ -cm column of DEAE-Sephacel  
5 equilibrated with 150 mM NaCl, 50 mM Tris-HCl, pH 8.0. After elution with 50 mM Tris-HCl (pH 8.0) containing 0.6 M NaCl, the glycanated glypican-1-Fc was bound to protein A-Sepharose beads and eluted with 0.1 M glycine, pH 3.0.

**10 Procedure**

1. Transfected into attached CHO stable cells with Lipofectamine 2000 in Opti-MEM 1. Overlay with DMEM media with 5% FBS after 4 hours.
2. Harvested after 3, 5 and 7 days incubation at 37°C.

**Cell Lysis/Protein Recovery****Procedure****15**

1. Centrifuged at 3000 rpm for 10 min and filter with 0.2  $\mu$ m pore size.

**Procedure**

1. Metal Affinity Chromatography – Pharmacia 50ml and 5 ml Metal Chelate – Running buffer 20 mM phosphate, pH 7.4, 0.5 M NaCl. Wash with 20mM, 50mM, and 100mM Imidazole. Elute with 500mM Imidazole.
2. HS Cation Exchange Chromatography—Poros HS 1.6 ml column—30 mM Tris-Cl, pH 8.0, 0.05% CHAPS. Elute with 0-2 M NaCl gradient.
3. Dialysis - @ 4°C using 3,500 MWCO against 20mM Tris-HCl, pH7.4 + 150mM NaCl.

**PROTEIN QUALITY CONTROL****Western Blot Procedure**

Antibody name, catalog # and supplier: Anti-V5-HRP Antibody, 46-0708, Invitrogen (Carlsbad, CA), S-protein HRP conjugate, 69047, Novagen (Madison, WI)

Antibody dilution buffer: PBS/5% milk/0.1% Tween-20

Wash buffer: PBS/0.1% Tween-20

Detection reagents: ECL (Amersham Biosciences Corp., Piscataway, NJ)

1. The blot was covered with antibody dilution buffer and incubated on a rocker for one hour at room temperature.
2. The blocking solution was replaced with antibody dilution buffer containing the appropriate amount of conjugate, and the blot was incubated on a rocking platform for one hour at room temperature.
3. The antibody solution was decanted, and the blot was washed quickly with two quick rinses of wash buffer. The blot was then covered with wash buffer and incubated on the rocking platform for five minutes, and the wash buffer was decanted. This process was repeated twice for a total of three five-minute washes.
4. The blot was developed using ECL reagents (Amersham Biosciences Corp., Piscataway, NJ) as per manufacturer instructions and luminescence was then digitized on a Kodak Image Sciences Imaging Station.

5

**Expression of CG50970-05 in stable CHO-K1 cells.**

A 1590 bp long BamHI-XhoI fragment containing the CG50970-05 sequence was subcloned into BamHI-XhoI digested pEE14.4Sec2 and pEE14.4SecFc3. The resulting plasmids are transfected into CHO-K1 cells using the LipofectaminePlus reagent following the manufacturer's instructions (Invitrogen/Gibco). The cell pellet and supernatant are harvested 72h post transfection and examined for CG50970-05 expression by Western blot (reducing conditions) using an anti-V5 antibody.

MSX resistant clones are selected using the GS system (Lonza Biologicals) The culture media in the selection process was: Glutamin-free DMEM (JRH), 10% dialyzed FBS, 1x GS supplement (JRH), 25uM MSX (JRH).

A high expressor clone, is selected for scale up in 10 LWave bioreactors. Two reactors were inoculated. 30 L conditioned media was collected from each reactors yielding batches 2 and 3.

The culture media was harvested 120h after inoculating the Wave bioreactor and examined for CG50970-05 expression by Western blot (reducing conditions) using an anti-V5 antibody

**Example E3. Growth factor mediated proliferation assays**

Several growth factors require the presence of heparan sulfate for high affinity binding to their tyrosine kinase receptors and therefore use HSPG's as coreceptors in their signaling. We determine whether it is possible to modulate responsiveness to

5 heparin-binding growth factors by altering CG50970 protein levels, either increasing them or decreasing them. Kleeff et al (J. Clin. Invest. Volume 102, Number 9, November 1998, 1662-1673) and Matsuda et al (Cancer Research 61, 5562-5569, July 15, 2001) used this approach to demonstrate the activity of Glypican-1.

Tumor cell lines with low level of CG50970 are transiently transfected with mature

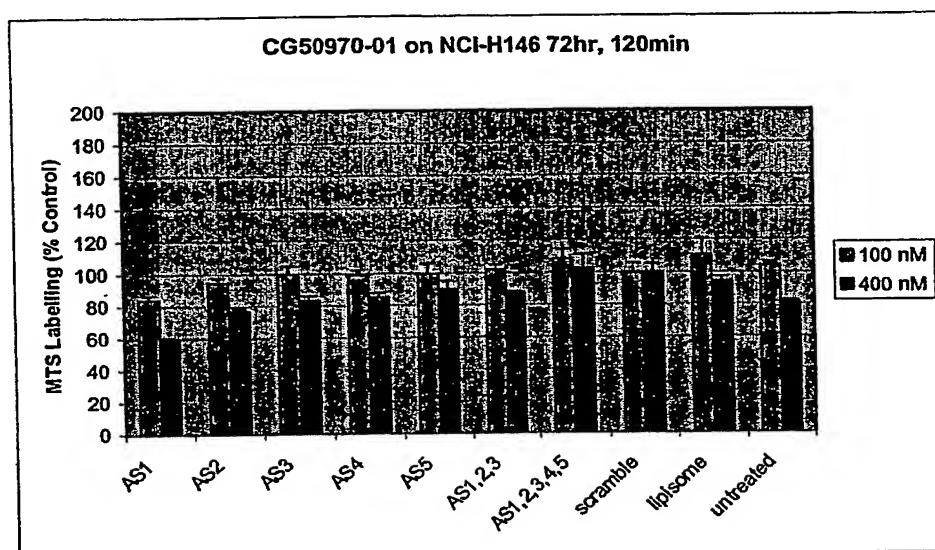
10 forms of CG50970, variants 06 and 07. The increase in expression of CG50970 is then monitored by western blot analysis. Next, the effects of growth factors on cell growth are determined during the 48-96 h interval after transfection, when CG50970 protein levels are maximally increased. Cells are treated with several growth factors like FGF2, HB-EGF. Cells expressing CG50970 will have a higher rate of proliferation in response to the growth

15 factors than control cells.

Tumor cell lines with high level of CG50970 are transiently transfected with antisense oligos directed against CG50970. The decrease in expression of CG50970 is then monitored by PCR-based methods. Next, the effects of growth factors on cell growth are determined during the 48-96 h interval after transfection, when CG50970 protein levels are

20 maximally decrease. Cells are treated with Fetal Bovine Serum or individual growth factors like FGF2, HB-EGF. As shown in table E3a below, cells treated with CG50970 antisense 1 and stimulated with with Fetal Bovine Serum have a lower rate of proliferation in response to the growth factors than control cells.

**Table E3a: Proliferation assay on NCI-H146 lung cancer cells treated with antisense oligos**



5                      Sequences of the antisense oligos, relative position and length that correspond to Table E3a.

AS1	ATGTCCGCGCTGCGACCTCT	1	20	0.0
AS2	ATGTCCGCGCTGCGACCTCT	1	20	0.0
AS3	CGGGAGCGAGGCAAAGGTCA	66	20	0.0
AS4	AACGACCGCCGCAGGCACCA	1137	20	0.0
AS5	GCTTGGACCTCGATAACGGG	1725	20	0.0

#### Example E4: Preparation of Antibodies that Bind CG50970

10                      As described above, inhibiting CG50970 activity has utility in cancer therapy and specifically in inhibiting lung and breast cancers. It is known in the art that antibodies that bind HSPGs factors like CG50970 can inhibit their activity in a process called neutralization. Specifically, neutralizing monoclonal antibodies that bind syndecan-3

15                      interfered with FGF-2 mitogenic action, but not that of insulin-like growth factor-1 or parathyroid hormone (Kirsch et al. J Biol Chem 2002 Nov 1;277(44):42171-7). Therefore production of polyclonal and monoclonal antibodies directed against CG50970 has utility in cancer therapy and specifically in inhibiting kidney, lung, melanomas and breast cancers. As opposed to VEGF, that is needed only for tumor induced endothelial cell growth and

survival, CG57094 is required for cell growth and survival both by endothelial and tumor cells, therefore inhibition of CG57094 activity could have a more pronounced therapeutic effect.

**Method:** Techniques for producing the antibodies are known in the art and are described, for example, in "Antibodies, a Laboratory Manual". Eds Harlow and Lane, Cold Spring Harbor publisher. Both rabbits and mice are suitable for the production of polyclonal antibodies, while mice are also suitable for the production of monoclonal antibodies. Mice where the human immunoglobulin genes have replaced the mouse immunoglobulin genes can be used to produce fully human monoclonal antibodies. These antibodies have better pharmaceutical characteristic, no or minimal antibody directed immune reactions that results in loss of therapeutic efficacy and have been shown to eradicate tumor in animal model (Yang XD, Jia XC, Corvalan JR, Wang P, Davis CG, Jakobovits A Eradication of established tumors by a fully human monoclonal antibody to the epidermal growth factor receptor without concomitant chemotherapy. Cancer Res 1999 Mar 15;59(6):1236-43).

#### 15      **Generation of rabbit polyclonal antibodies**

Rabbit are immunized with the immunogen emulsified in complete Freund's adjuvant and injected subcutaneously or intraperitoneally or intramuscular in an amount from 50-1000 micrograms. The immunized rabbits are then boosted 10 to 12 days later with additional immunogen emulsified in the selected adjuvant. Thereafter, for several weeks, the rabbits might also be boosted with additional immunization injections. Serum samples may be periodically obtained from the rabbit by bleeding of the ear for testing ELISA assays to detect the antibodies.

#### 25      **Generation of human monoclonal antibodies**

Fully human monoclonal antibodies (MAb), direct against CG50970-05 are generated from human antibody-producing XenoMouse strains engineered to be deficient in mouse antibody production and to contain the majority of the human antibody gene repertoire on megabase-sized fragments from the human heavy and kappa light chain loci as previously described in Yang et al. (Eradication of established tumors by a fully human monoclonal antibody to the epidermal growth factor receptor without concomitant chemotherapy. Cancer Res 1999 Mar 15;59(6):1236-43).

Elisa assay is then used to determine the specificity of the antibodies.

**Example E5: ELISA Protocol to determine binding of the antibodies****Solution Preparation**

- Coating Buffer (0.1M Carbonate, pH9.5)
- 8.4 g. NaHCO<sub>3</sub>, 3.56g. Na<sub>2</sub>CO<sub>3</sub>, pH to 9.5, and dilute to 1 L. with ddH<sub>2</sub>O

5

**ASSAY DILUENT**

- Pharmingen #26411E

**PROTOCOL**

- Coat a 96-well high protein binding ELISA plate (Corning Costar #3590) with 50  
10 ul. of protein at a concentration of 5ug/mL. in coating buffer overnight at 4 degrees.
- Following day wash the cells 5X 200-300 ul. of 0.5% Tween-20 in PBS.
- Block plates with 200ul. of assay diluent for at least 1 hour at room temperature.
- Dilute antibodies in assay diluent.
- Wash plate as in step 2.
- Add 50ul. of each antibody dilution to the proper wells for at least 2 hours at room  
15 temp.
- Wash plate as in step 2.
- Add 50ul. of secondary antibody and incubate for 1 hour at room temp.
- Wash plate as in step 2.
- Develop assay with 100ul. of TMB substrate solution/well. (1:1 ratio of solution  
20 A+B) (Pharmingen #2642KK)
- Stop reaction with 50ul. sulfuric acid

Read plate at 450nm with a correction of 550nm.

**Example 6: Identification of CG50970 neutralizing antibodies**

- 25 As mentioned above, proteoglycans like CG50970 have affinity for glycosaminoglycan (GAG)-binding proteins like laminin-1 and midkine. Specifically, Herndon et al (Glycobiology 1999 Feb;9(2):143-55) have previously shown that rat glypican-2 has an high affinity for laminin-1, while Kurosawa et al. (Glycoconj J 2001 Jun;18(6):499-507) have shown that rat glypican-2 has an high affinity for midkine.

As previously discussed, the identification of antibodies, preferably fully human monoclonal antibodies that bind to CG50970 and neutralize its activity, limiting or abolishing its ability to bind to glycosaminoglycan (GAG)-binding proteins like laminin-1 and midkine, would be very beneficial because these antibodies will have therapeutic effect against tumors, specifically against lung and breast cancers. To determine whether an antibody can neutralize CG50970 activity, various amounts of such antibody are added to the Receptor-ligand Elisa assay as described in the method below.

**Receptor-ligand Elisa assay Protocol** – 96-well plates (Corning Costar, catalog no. 9018) are coated overnight with the laminin-1 (BT-276 from BTI website at [btiinc.com/page/catal.html#Laminin](http://btiinc.com/page/catal.html#Laminin)) or midkine ( 258-MD from R&D system) at a saturating concentration of in phosphate-buffered saline. After removing the unbound protein by washing with TBST (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.1% Tween 20), the wells were blocked with 10% fetal bovine serum in TBST for 2 h and then incubated for 18 h at room temperature with varying concentrations of glycanated CG50970-05-Fc in phosphate-buffered saline in the presence or absence of various amounts of monoclonal antibodies that bind to CG50970. The CG50970-05-Fc bound to laminin-1 proteins was detected using a biotinylated anti-human Fc antibody (Jackson ImmunoResearch Laboratories, Inc.; 1:250,000 in TBST, for 2 h) followed by incubation for 20 min with horseradish peroxidase-conjugated streptavidin (1:20,000 in TBST). The colorimetric reaction product from the *o*-phenylenediamine substrate was measured at 450 nm using a Dynatech MRX ELISA plate reader. Nonspecific binding was calculated as the binding of glypican-2-Fc to wells coated with 100 µg of bovine serum albumin.

Antibodies identified with this assay are then tested at various concentrations in the growth factor mediated proliferation assay described in example 4 to determine whether they can inhibit cellular proliferation.

Antibody that can neutralize the CG50970-05-Fc biochemical activity and have anti-proliferative activity can be useful as therapeutic agents.

#### **Example 7: Quantification of membrane bound CG50970 by Flow Cytometry**

CG50970 is a type 1 membrane protein, therefore Mabs binding to this protein could be able to stain the membrane of cells expressing CG50970 in a Flow Cytometry

assay (FACS). It is known in the art that not all antibodies that recognize a recombinant protein in Elisa or IHC assays will also work in FACS. At the same time those antibodies that do are preferred because they have a higher chance to recognize the antigen in-vivo in patients and therefore have a potential use as therapeutic or ex-vivo diagnostic agents. We  
 5 therefore set-up a FACS assay using cell lines that express CG50970, like lung ca.ncer NCI-H146, NCI-H526 or breast cancer BT 549 and one that express it at much lower level, lung ca.ncer HOP-62 and breast cancer T47D.

#### Flow Cytometry Protocol for Adherent Cells

(ver.1) 11-25-02 KT

10

1. Wash cells with 1x PBS (Ca and Mg free) twice.
2. Add Versene and incubate at 37°C until the cells detach.
3. Count cells. Use <1 million cells per assay tube.
4. Wash the cells twice with ice-cold FACS buffer.

15

5. Resuspend cells in 100 ul of ice-cold FACS buffer. Mix.
6. Add primary mAb. Incubate on ice for 30 min.
7. Wash cells 2-3 times with 1 ml of ice-cold FACS buffer.
8. Resuspend cells in 100 ul ice-cold FACS buffer. Mix.
9. Add secondary (conjugated) mAb. Incubate on ice for 30 min with a cover.

20

10. Wash cells 2-3 times with 1 ml of ice-cold FACS buffer.
11. Fix cells with 0.5-1 ml of 1 % formaldehyde (in PBS) and analyze by Flow Cytometry.

#### FACS buffer:

25

0.01 M HEPES (pH 7.4)

0.15 M NaCl

----- (may be substituted by PBS)

0.1% NaN<sub>3</sub>

4 % FBS

30

#### **Example 8: Preparing and testing of chemotherapy and radioimmunoconjugated antibodies**

Cytotoxic chemotherapy or radiotherapy of cancer is limited by serious, sometimes  
 35 life-threatening, side effects that arise from toxicities to sensitive normal cells because the therapies are not selective for malignant cells. There therefore the need to improve the

selectivity. One strategy is to couple the therapeutics to antibodies that recognize tumour-associated antigens. This increases the exposure of the malignant cells, and reduces the exposure of normal cells, to the ligand-targeted therapeutics (reviewed in Allen Ligand-targeted therapeutics in anticancer therapy. Nat Rev Cancer 2002

5 Oct;2(10):750-63)

CG56972-03 is one of these tumour-associated antigen, as shown by its specific expression on cellular membranes of tumor cells by FACS and IHC.

Therefore the fully human monoclonal antibodies direct against CG50970-05  
10 disclosed in this application could be coupled to cytotoxic chemotherapeutic agents or radiotherapeutic agents to generate anti-tumor therapeutics.

Depending on the intended use of the antibody, i.e., as a diagnostic or therapeutic reagent, radiolabels are known in the art and have been used for similar purposes. For  
15 instance, radionuclides which have been used in clinical diagnosis include <sup>131</sup>I, <sup>125</sup>I, <sup>123</sup>I, <sup>99</sup>Tc, <sup>67</sup>Ga, as well as <sup>111</sup>In. Antibodies have also been labeled with a variety of radionuclides for potential use in targeted immunotherapy (Peirersz et al. (1987) The use of monoclonal antibody conjugates for the diagnosis and treatment of cancer. Immunol. Cell Biol65: 111-125). These radionuclides include <sup>188</sup>Re and <sup>186</sup>Re as well as <sup>90</sup>Y, and to a lesser extent <sup>199</sup>Au and <sup>67</sup>Cu.  
20 I-(131) has also been used for therapeutic purposes. U.S. Pat. No. 5,460,785 provides a listing of such radioisotopes and is herein incorporated by reference.

Patents relating to radiotherapeutic chelators and chelator conjugates are known in the art. For instance, U.S. Pat. No. 4,831,175 of Gansow is directed to polysubstituted  
25 diethylenetriaminepentaacetic acid chelates and protein conjugates containing the same, and methods for their preparation. U.S. Pat. Nos. 5,099,069, 5,246,692, 5,286,850, and 5,124,471 of Gansow also relate to polysubstituted DTPA chelates. These patents are incorporated herein in their entirety.

30 Cytotoxic chemotherapy are known in the art and have been used for similar purposes. For instance . U.S. Pat. No 6,441,163 describes the process for the production of cytotoxic conjugates of maytansinoids and antibodies. The anti-tumro activity of a new tubulin polymerization inhibitor, auristatin PE, is know in the art (Mohammad et al. Int J Oncol 1999 Aug;15(2):367-72).

Once these conjugates of chemotherapy or radiolabels and antibodies is made, it is tested for its cytotoxic activity on CG50970-05 positive cells, using methods know in the art like by MTS, Cell counts and clonogenic assays.

5

### OTHER EMBODIMENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is

5 contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. The choice of nucleic acid starting material, clone of interest, or library type is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein. Other aspects, advantages, and

10 modifications considered to be within the scope of the following claims. The claims presented are representative of the inventions disclosed herein. Other, unclaimed inventions are also contemplated. Applicants reserve the right to pursue such inventions in later claims.

## CLAIMS

What is claimed is:

1. An isolated polypeptide comprising the mature form of an amino acid sequenced selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52.
2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52.
3. An isolated polypeptide comprising an amino acid sequence which is at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52.
4. An isolated polypeptide, wherein the polypeptide comprises an amino acid sequence comprising one or more conservative substitutions in the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52.
5. The polypeptide of claim 1 wherein said polypeptide is naturally occurring.
6. A composition comprising the polypeptide of claim 1 and a carrier.
7. A kit comprising, in one or more containers, the composition of claim 6.
8. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein the therapeutic comprises the polypeptide of claim 1.
9. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

- (a) providing said sample;
- (b) introducing said sample to an antibody that binds immunospecifically to the polypeptide; and
- (c) determining the presence or amount of antibody bound to said polypeptide, thereby determining the presence or amount of polypeptide in said sample.

10. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the polypeptide of claim 1 in a first mammalian subject, the method comprising:

- a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and
- b) comparing the expression of said polypeptide in the sample of step (a) to the expression of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease,

wherein an alteration in the level of expression of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

11. A method of identifying an agent that binds to the polypeptide of claim 1, the method comprising:

- (a) introducing said polypeptide to said agent; and
- (b) determining whether said agent binds to said polypeptide.

12. The method of claim 11 wherein the agent is a cellular receptor or a downstream effector.

13. A method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of the polypeptide of claim 1, the method comprising:

- (a) providing a cell expressing the polypeptide of claim 1 and having a property or function ascribable to the polypeptide;
- (b) contacting the cell with a composition comprising a candidate substance; and

- (c) determining whether the substance alters the property or function ascribable to the polypeptide;

whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition in the absence of the substance, the substance is identified as a potential therapeutic agent.

14. A method for screening for a modulator of activity of or of latency or predisposition to a pathology associated with the polypeptide of claim 1, said method comprising:

- (a) administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of claim 1, wherein said test animal recombinantly expresses the polypeptide of claim 1;
- (b) measuring the activity of said polypeptide in said test animal after administering the compound of step (a); and
- (c) comparing the activity of said polypeptide in said test animal with the activity of said polypeptide in a control animal not administered said polypeptide, wherein a change in the activity of said polypeptide in said test animal relative to said control animal indicates the test compound is a modulator activity of or latency or predisposition to, a pathology associated with the polypeptide of claim 1.

15. The method of claim 14, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.

16. A method for modulating the activity of the polypeptide of claim 1, the method comprising contacting a cell sample expressing the polypeptide of claim 1 with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.

17. A method of treating or preventing a pathology associated with the polypeptide of claim 1, the method comprising administering the polypeptide of claim 1 to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject.

18. The method of claim 17, wherein the subject is a human.
19. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52 or a biologically active fragment thereof.
20. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52.
21. The nucleic acid molecule of claim 20, wherein the nucleic acid molecule is naturally occurring.
22. A nucleic acid molecule, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 52.
23. An isolated nucleic acid molecule encoding the mature form of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 52.
24. An isolated nucleic acid molecule comprising a nucleic acid selected from the group consisting of 2n-1, wherein n is an integer between 1 and 52.
25. The nucleic acid molecule of claim 20, wherein said nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 52, or a complement of said nucleotide sequence.
26. A vector comprising the nucleic acid molecule of claim 20.

27. The vector of claim 26, further comprising a promoter operably linked to said nucleic acid molecule.
28. A cell comprising the vector of claim 26.
29. An antibody that immunospecifically binds to the polypeptide of claim 1.
30. The antibody of claim 29, wherein the antibody is a monoclonal antibody.
31. The antibody of claim 29, wherein the antibody is a humanized antibody.
32. A method for determining the presence or amount of the nucleic acid molecule of claim 20 in a sample, the method comprising:
- (a) providing said sample;
  - (b) introducing said sample to a probe that binds to said nucleic acid molecule; and
  - (c) determining the presence or amount of said probe bound to said nucleic acid molecule,
- thereby determining the presence or amount of the nucleic acid molecule in said sample.
33. The method of claim 32 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.
34. The method of claim 33 wherein the cell or tissue type is cancerous.
35. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the nucleic acid molecule of claim 20 in a first mammalian subject, the method comprising:
- a) measuring the level of expression of the nucleic acid in a sample from the first mammalian subject; and
  - b) comparing the level of expression of said nucleic acid in the sample of step (a) to the level of expression of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease;

wherein an alteration in the level of expression of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

36. A method of producing the polypeptide of claim 1, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52.

37. The method of claim 36 wherein the cell is a bacterial cell.

38. The method of claim 36 wherein the cell is an insect cell.

39. The method of claim 36 wherein the cell is a yeast cell.

40. The method of claim 36 wherein the cell is a mammalian cell.

41. A method of producing the polypeptide of claim 2, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 52.

42. The method of claim 41 wherein the cell is a bacterial cell.

43. The method of claim 41 wherein the cell is an insect cell.

44. The method of claim 41 wherein the cell is a yeast cell.

45. The method of claim 41 wherein the cell is a mammalian cell.

THIS PAGE BLANK (USPTO)